

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**



CC The present invention is related to fluorescent apophytochrome-bilin  
CC conjugates, known as phytofluors. An apoprotein known as Cph2 from  
CC *Synechocystis* species is used as the apophytochrome and the bilin is  
CC preferably phycoerythrobilin. The phytofluors are useful as fluorescent  
CC markers for biological research. The phytofluors have a long wavelength  
CC absorption maxima, a high molar absorption coefficient and the  
CC recombinant apoproteins can spontaneously assemble with a variety of  
CC bilin chromophore precursors. The present sequence is a phytochrome  
CC related protein from *Synechocystis* sp.

XX Sequence 1276 AA;

Query Match 100.0%; Score 6614; DB 21; Length 1276;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNPNLSLDFLRNNTKRRALTLRETLQVYEAKRIFLGVDRVATYFASGSGEVLAE 60  
DB 1 mnpnlsldflrnntkrraltlretrlqvyeakriflgvdrvatyf asgsg evlae 60  
QY 61 AVNRALPSLGLHPVPDIPPOAREELGNORKMIADVARRKKSHELSGRISPTESHN 120  
DB 61 avnraalpsllglhvpdippoareelgnorkmiadvarrkkshe lsgrisp teshn 120  
QY 121 GHYTTVDSCHIQVLLAMGVSSLVTVPNQDOQLMGIMAVHHSKPRPTEOEMETWALISK 180  
DB 121 ghyttvdsch iqvllamgvsslv tvpnqdqlwgmavhhs kprpteoem etwalisk 180  
QY 181 EVSLAITQSLSROYHQOQVEALYQRLVETVAYGDRPRTVOVALLEVGVGADEVGAVL 240  
DB 181 evslaitqsls royhqoqv eal yqrlvetvay gdrprtvoval levvgvga devgavl 240  
QY 241 YIAPDLTGSVAOHQOMNRFDMGNMLETSLMOELRGOPSAEMPMVAOSTWEKRPPT 300  
DB 241 yiapdl tgsvaohqom nrfdmgnmlet slmoelrgop saempmva ostwek rpppt 300  
QY 301 SVAPLPTNCVPHGYTGLGELEQRSDWIAPPESLSAENFOSFLYPLAADQGVGSLILLR 360  
DB 301 svaplptncv phgytglgeleqrsdwia ppe ssaenfos flyplaa dqvgvgs llllr 360  
QY 361 KEKSLVKRMWAKRGIDRNIIIRLSFEAMETOKIVPTWNSERKLAQVASTQLYMATTO 420  
DB 361 kek slvkrmwakrgid rniir lsfeamet okivptwn serk laqvastq ly matto 420  
QY 421 QFVRLTIQQTAYDPLTOLPMNIIFNRQLTALDLATLEGKMWGLVYAMDREKXINESF 480  
DB 421 qfvrlti qqtaydpl t olpmniifnrql taldlat legkmwglv yamdre kxinesf 480  
QY 481 GHKTGDLLOEVADRLNOKLSPLAAYSPILSRWHGDFITLLTQISDNQEMIPLCERLLS 540  
DB 481 ghktgd lloevadrl noklsplaay spilsr whgdfitll t qisdnq emiplcerlls 540  
QY 541 TFOEPFLQOGPIYLTASMGISTAPYDGETAESLKFPEILATRAKCOGKNTYOFERYPD 600  
DB 541 tfoepfl qogpiy l tasmgist apydg etaes l kfp eilatra kogknt yofery pd 600  
QY 601 SAPMLDRLTLESDDLROATLNOEFLYRPOVALDPTGKLGEVALVWQHPRLGVAAPVYF 660  
DB 601 sapmldr ltlesddl roatlnoe flyrpo valdptgkl gevalvwqh prlgva apvyf 660  
QY 661 IPLAEELGLINHLGOWVLETACATHOHFFRETGRRLRMANVISARQOFDEKMLNSVLECL 720  
DB 661 iplaeel glinhlgowv letacath ohffret grrlrman visarq ofdekmlnsvlecl 720  
QY 721 KRTGMPEBDELEITESTLAMEDIKGVVLLHRLREBQOVAIDFGTGYSSLSILIKQLPI 780  
DB 721 krtgmpeb deleitestlamedikgvv llhrlre bqvaid fgtgyssls ilikqlpi 780  
QY 781 HRLTIDSFVNDLNEGADTAIIQYVIDLANGLNEFTAAEISEEAOQRIQKMGCHIQO 840  
DB 781 hrltid sfvndl negadta i iqyvidlang lneftaa eiseeao qriqkmgch iqo 840

QY 841 GYFLTRPLPAAAMWTLYXPQILDGFPPLPKVALPETETEGAGGNVGRPLNSLRE 900  
DB 841 gyfltrpl paaamwtly xpqildg fpp lpkval petet ega gnvgrplnslre 900  
QY 901 NPWTEKLDHYVLKRLDOORNVKEKLVLIKANKIRASLINDILYSTYEVROFLNDRV 960  
DB 901 npwtekl dhyvlk rldoor nvkek lvlikan kirasl indil ystyev roflndrv 960  
QY 961 VLFKNSQMSQVWVESHNDCRSITINDEIDPCFKGHYRLRYRGVRAVSDIEKADLA 1020  
DB 961 vlfknsq msqv wvesh ndcrs itin deidpc fkg hylrly rgvra vsdiek adla 1020  
QY 1021 DCHKELRHYOKANLVVYVFNENLWGLLAHECKPRRYQOEDDLQIMELATQVALAI 1080  
DB 1021 dchkelr hyokanlv vyvfn enlwgl lahe ckpr ryqo eddlq imelat qvalai 1080  
QY 1081 HOGELVEQLERANTRLQOISSLDALTOYGNRYLFDSTLEREMORLQRTREPLALLCOVD 1140  
DB 1081 hogelve qlerant r lqoissld altoy gnryl fdstler emorl qrtrepl allcovd 1140  
QY 1141 FFKGFNDYGHPRADRCCLKIADAMAKYAKRPTDLVARYGGEPAIILSETSLGAINVT 1200  
DB 1141 ffk gfn dygh pradrc clk iadama kyakrptdl var yggepai ilsetsl gainvt 1200  
QY 1201 EALQYEVANLAIPTVSGTGHVTLSIGIAYVTPERHINPNALVKAADLALYAKAKGRNQ 1260  
DB 1201 ealqyev anlai p tvsgt ghv t l sigiay vtp erhin pnalvka adlaly akakgrnq 1260  
QY 1261 WLAYEGSOLPHVDGEV 1276  
DB 1261 wlayeg sq lphvd gev 1276

## RESULT 2

ID AAB26600 standard; peptide: 212 AA.  
AC AAB26600;  
XX 01-FEB-2001 (first entry)  
DE

XX *Synechocystis* sp phytochrome-related gene Cph2-N197 peptide.  
XX phytochrome; fluorescent apophytochrome; bilin; Cph.  
XX *Synechocystis* sp.  
XX W0200056355-A1.  
XX 28-SEP-2000.  
XX 14-MAR-2000; 2000WC-US06607.  
XX 19-MAR-1999; 99US-0272809.  
XX (REGC ) UNIV CALIFORNIA.  
XX

XX Lagarias JC;  
XX WPI: 2000-602195/57.  
XX

PT Fluorescent apophytochrome-bilin conjugates, e.g. Cph2 (derived from  
PT *Synechocystis* species) and phycoerythrobilin conjugate, useful as  
PT fluorescent markers for biological research -

PS Claim 5; Fig 3; 52pp; English.

CC The present invention is related to fluorescent apophytochrome-bilin  
CC conjugates, known as phytofluors. An apoprotein known as Cph2 from  
CC *Synechocystis* species is used as the apophytochrome and the bilin is  
CC preferably phycoerythrobilin. The phytofluors are useful as fluorescent  
CC markers for biological research. The phytofluors have a long wavelength  
CC absorption maxima, a high molar absorption coefficient and the



CC recombinant apoproteins can spontaneously assemble with a variety of  
CC bilin chromophore precursors. The present sequence is a phytochrome  
CC related peptide from *Synechocystis* sp.  
XX  
SQ Sequence 212 AA;

Query Match 15.2%; Score 1003; DB 21; Length 212;  
Best Local Similarity 100.0%; Pred. No. 3,1e-78;  
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNPRSLDEPFRVYINKFRALTLRETLQVVEARIFLGVDRKITYFASDGEVLAIE 60  
Db 1 mnprrsledefrnyvinkfraltlreqlqvveeariflgvdrkityfascgsgevlae 60  
OY 61 AVNRAALPSLGLHPVEDIPPOAREELGNORKMIADVARRKKSHELSGRISPTESHN 120  
Db 61 avnraalpsllglhfpvedippargeelgnrkmladvahrkkshelesgrispteshn 120  
OY 121 GHYTVDSCHIQVLLAMGVLSLTVPVMDQOLMGIMAVHNSKRRFRTEQEWETMALISK 180  
Db 121 ghytvdschqyllamgvlsltvpvmdqqlwylimavhnskprfrteqewetmalisk 180  
OY 181 EVSLAITQSOLSRQVHQ 197  
Db 181 evslaitqsglsrqvqh 197

RESULT 3  
ID AAR38153 standard; Protein; 752 AA.  
XX AAR38153;

XX 13-OCT-1993 (first entry)

DE Acetobacter diuanylate phosphodiesterase PDEA2.

KW Cyclic diuanylate; diuanylate phosphodiesterase;

XX diuanylate cyclase; cellulose production; cdg2 operon.

OS Acetobacter xylinum.

XX  
FH Key Location/Qualifiers

FT Misc-difference 77 /note= "Val deduced from GC"

FT Misc-difference 212 /note= "Met deduced from TAG"

FT Misc-difference 215 /note= "Arg deduced from GC"

FT Misc-difference 232 /note= "Met deduced from ATC"

FT Misc-difference 233 /note= "Leu deduced from GTC"

FT Misc-difference 247 /note= "Gly deduced from GC"

FT Misc-difference 271 /note= "Gln deduced from CAG"

FT Misc-difference 275 /note= "Val deduced from GCTG"

FT Misc-difference 388 /note= "Asp deduced from GAA"

FT Misc-difference 538 /note= "Ala deduced from CC"

XX WO9311244-A.

XX 10-JUN-1993.

XX 14-OCT-1992; 92WO-US08756.

XX 29-NOV-1991; 91US-0800218.

PA (WEYE ) WEYERHAEUSER CO.

XX Ben-Bassat A, Benziman M, Calhoon RD, Gelfand DH;

PI Tal R, Wong HC;

XX WP: 1993-197062/24.

DR N-PSDB; AA043661.

PT Polynucleotide sequence from Acetobacter cdg operon - encodes  
PT cyclic di:guanosine mono:phosphate degradation enzymes e.g.  
PT 3-phosphodiesterase isozyme

PS Claim 5; Page 80-83; 98pp; English.

XX The amino acid sequence of protein PDEA2 was deduced from the 1st.  
CC open reading frame of the cdg2 operon. The protein is a diuanylate  
CC phosphodiesterase A, i.e. it enzymatically cleaves a single  
CC phosphodiester bond in c-di-GMP to yield the linear dimer pGpG.  
XX See also AAR38154.

SQ Sequence 752 AA;

Query Match 10.4%; Score 689; DB 14; Length 752;  
Best Local Similarity 26.2%; Pred. No. 4.4e-50;  
Matches 228; Conservative 131; Mismatches 322; Indels 188; Gaps 19;

OY 19 HRALTLRETLQVVEARIFLGVDRKITYKFA SDGGEVLA EAVNRALPS-----L 70

Db 11 hgavgradvlefvtvdqpl-----sadvllsamqadlpmvndehlli 54

OY 71 LGHPVEDIPPOAREELGNORKMIADVARRKKSHELS-----GRISPT-----E 117

Db 55 iffnaaekiwgcreeevmgnvsclypepdrtdndylnnretgygrlvgsrvefr 114

OY 118 HSNCHY-----TVDSCHIQVLLAMGVLSLTVPVMDQOLMGIMAVHNSKRRFR 168

Db 115 rangeyvcgelstsrvgyndcklyla-----vmkd-----vteqsrqkll 157

OY 169 EEWETMALSKREVSLATITQSOLSRQVHQVOEALVORLETTVAOYGDREPTWOYALET 228

Db 158 vlqndvlgalasmsl-----gevgdlldlcradsfvp----- 189

OY 229 VGOAVEADGAVLYTAPDLTGSVAOHTQWNLFDMGNNLETSLKQELMGRGSAAMEPMAA 288

Db 190 -----gavgalmlldpsrrllsvs-----aspsmpkryaa 219

OY 289 VQSTWEKRRPFTSVA--PLPPTNCVPHGVTIGELBQSRDWTAPRPSLAEMFQSLTVPL 346

Db 220 ldsmgllrpeqlrmlraapagsnlvvdgaa-----slarslglrccssllsr 268

OY 347 AADQWVGSLLLRKREKSLVKNMAGKRIGIDRRNLLPLRSPFAWEETQKLVPTWNRSEKRL 406

Db 269 tggymvfyalyllrgdeadl-----awag----- 291

OY 407 AOVASTOL--YMAITQOFVTRLTQOTADPDLQLPWIIIFNRQTLALDALY---EG 460

Db 292 -rvvstmpfcalaiegsetrqhlaqnsfslcgl-----lntslmhterllimggd 345

OY 461 KMGCVLVIAMRFRKRNESFGHKTGDDGLQEVADRLNKLSPRLAYSLSLRWNGDGTI 520

Db 346 sqfslmvdldfrldlnalghvnaadrfllelgrtrln---lvkddyivrsrggdefl 401

OY 521 LITQISDNQEMIRPLCERLSTFQEPFFLOGQRIYLTASMSGISTAPYDSETSLSLKFAEI 580

Db 402 vvpccs-herckfaenlinalaiairplygenttiscvsgistfrangpdesllshada 460

OY 581 ALTRAKCGKNTYQFRRPQDSAPMLDRLLTSDLRQALTNGEVLVYQFOVALDTGKLLG 640

Db 461 atrqakedgrylfrfaegeknvagdrilvgsalrdsiskgmllnynpqrvecltgg11g 520

OY 641 VEALVRMQRRLGOVARPVFTRLADELGLNHLQWLETCALNHNQFFRTGRLMAV 700

Key	Location/Qualifiers	PH
QY	115 PHEHNGHVTYTD-----SCHIOYLLAMGVL-----SLTPRVNQQDOIM--G 150	
Db	59 PLLHQAghdaIvetersgrsshnrtIvgtselevelrfsrdsgeyIgeIsIskVvddkrlfImg 118	
QY	156 IM--VHHHSKPRPRFEQEMETMALSKSVSLAITSQSLSRQVHOQOYVQALVQRLETTVA 213	
Db	119 ymkhvtuesqgrK-----IILLqndvIqalasdmI-----qdvadllcrrvesfv- 164	
QY	214 QYGRPEFTWQVLALETVGQAVEADGAVLYIAEDLTGSAQHGYMULREDPWMMLETSLMOE 273	
Db	165 -----pgt-----vaVnlllprdqqlvIs---spLlPKryraSe---slyvsaselek 208	
QY	274 LMRGQPSAMERMAVOSTWKEPRRFTSVAPLRPTNCVPHQYTLTGCELEQRSDWIAPESL 333	
Db	209 l-----rvpKhaktrmwdsytsIgsIsIqlgcfcIcpvstrsqgrK----- 249	
QY	334 SAENFQSLVLPVLAADQWGSLLILRKESKLVHMKAGRGIDRRNILLPSFAMEBETQ 399	
Db	250 -----gfaIysrdsdgrntwP-grIvdsclpIcalaIeq----- 283	
QY	394 KLVPVWNRSEKRLQVASTQ-----LWMAITQGVYTRILITQGTAVDPLTQLPNNIIRNQ 444	
Db	284 -----natgeIsInIuaIafsltgllnrsvshvKiegnIsIsqdg-----nrg 322	
QY	449 LTLALLDALYEGKMGVGLVVIAMDRKRIINESFGHKGTGDGLQEVADRLNOKLSPLAASP 508	
Db	325 faIrnId-----IdfrIdnIdalghvnyadqqlfIeIaIarIs---Iakedy 366	
QY	509 LLSRHHGCGFTILLTQISDNDQMPLRCRLSLSTQEPFFLQGPFIYLTASGISIATFYDG 565	
Db	367 vIstrsggeIvfvvvpd-cphkeactdIaehllasmtcmpmgIqnclIsIscIsIgstpydng 423	
QY	569 ETAEISLTFEAELATLTRAQCGCKNTYQOEFRRPDSAPMLDRLTLESBDLROALTNQDEFVLFQ 628	
Db	426 pdsesllstadvallrqakeDgrYfrfIenIknqvaqdrlvIgsalIstslakgmIlnIhyg 485	
QY	629 POVALDNGKLLGVEALVWQHPRLGQVAAADVFIPLAEGLINHLGQWVLETCATQHOF 688	
Db	486 pqvrtchIeIsgvealsrwhphrIghIfrsfrIaVaeegIealgrwsllaeacrqIvKw 543	
QY	689 FRETRRRLRMVNISARQFODEKMLNSVLECLTKTGMPREDLELEIETESLMEDIKSTNV 748	
Db	546 dtdqIhIprvtaVnIasvInrraIpenIaallkohnIkpstllveIeasvnmndsrItee 605	
QY	749 LLHRLREBGVOVAIDDFGTGYSSLSILKQDPIHRLKIDKSFVNDLNEGADTALLIQYVID 808	
Db	606 vIqstInIgcglsmddfgItyssIsrltrIprIeIkIdrIsIndfendhnaqavtmavIg 665	
QY	809 LANGINLETVADGIESEKQIOLRQKMGCHLGQGFILTRPLPAEKMMYIYVLPQ-ILDFGR 867	
Db	666 IqsrIgmIvuvIeegveteqgrIdlekIIndcmvgyllIakrIplapddfexkmwtrhngItrgmIp 723	
QY	868 TPPLPK 873	
Db	726 aaapaK 731	
RESULT	5	
AAR38150		
ID	AAR38150 standard; Protein: 765 AA.	
XX		
AC	AAR38150;	
XX		
XX	13-OCT-1993 (first entry)	
DE	Acetobacter diuanylate phosphodiesterase PDEA1.	
XX		
KW	Cyclic diuanylate: diuanylate phosphodiesterase;	
XX	diuanylate cyclase; cellulose production; cgl operon.	
XX	Acetobacter xylinum.	
XX		
XX		



## RESULT 7

AAR41019 standard; Protein: 340 AA.

ID AAR41019;

AC AAR41019;

DT 25-MAR-1994 (first entry)

DE Insecticidal protein gene ORF-1 prod.

XX Caulobacter; plasmid; insecticidal protein; Bacillus thuringiensis;

KW Bacillus sphaericus; larva; mosquito; Culex; Anopheles; Psorophoa;

KM Mansonla; Aedes.

XX Bacillus sphaericus strain SSI-1.

OS JPO5211866-A.

XX 24-AUG-1993.

XX 05-JUN-1991; 91JP-0160963.

XX 06-JUN-1990; 90JP-0148444.

XX (SIIM-) SILMARAN SO TANABAL.

XX WPI: 1993-298916/38.

XX N-PSDB: AAQ48715.

XX Expression of insecticidal protein - by transforming Caulobacter

PT with plasmid contg. gene coding for insecticidal protein

PS Disclosure; Page 16-21; 27pp; Japanese.

XX Caulobacter transformed with a plasmid contg. a gene encoding

CC insecticidal protein derived from Bacillus thuringiensis or

CC Bacillus sphaericus will proliferate in aq. environment.

CC They may be consumed by larvae of mosquitoes and are lethal to

CC Culex, Anopheles, Psorophoa, Mansonla and Aedes.

XX Sequence 340 AA;

Query Match 9.6%; Score 635; DB 14; Length 340;

Best Local Similarity 39.9%; Pred. No. 5.2e-46;

Matches 130; Conservative 77; Mismatches 115; Indels 4; Gaps 1;

XX 536 ERLS---TQDEPFLQGPPIYLTLASMGISTAPYDGTASLKFALTRAKCOGKN 591

DB 6 ekmlsslrdsfvqpfllngqiltfvqmslgiasypidgytgeellkhadlanykakel9gn 65

XX 592 TYOFYRPQDSAPMLDRLTLSDLRQALTNQEFVLYFQPOVALDYGKLGVEALVRWQHPR 651

DB 66 nhrffdekmmelavlkdqiermitlalernefsvhyqdiaetgkiryfealvrvkspe 125

XX 652 LGQVAPVFIPLAEELGILNHLGQVLETCATQHOFRETRGRRLMAVNISARQFODEK 711

DB 126 lglvspeffipiaektgiltqldewwmyqaciknvelqhgfygpfllmsvnlsalqlgrad 185

XX 712 WLNSVLECLKRTGMPPELEITESLMEEDIKGTIVLLHLRREGVOVAIDDEGTGYSS 771

DB 186 fvdvkvqllnctkmkpenheleiteslivesfessicllrklngvklagddfgtgyss 245

XX 772 LSIKOLPIHRLKIDKSVNVLNMGATATIOYVIDLANGLNETVAGEISEAQOLRL 831

DB 246 lnytlilpibhklkdrslqgmtsataektliesihlahkngdvaaegvetekeyyll 305

XX 832 QKMGCHLGQGYFLTRPLPAEAMMTYL 857

DB 306 kemncdfvggyfrrpsvssdlvecl 331

## RESULT 8

AAR75409 standard; Protein: 340 AA.

ID AAR75409;

AC AAR75409;

DT 20-NOV-1995 (first entry)

DE B. sphaericus mtx toxin-associated protein.

XX Biological control agent; mosquitoicidal toxin; insecticide;

KW mosquito; Culex; Aedes; Anopheles.

XX Bacillus sphaericus.

OS W09515383-A.

XX 08-JUN-1995.

XX 30-NOV-1994; 94WO-GB02628.

XX 30-NOV-1993; 93GB-0024529.

XX (UTSI-) UNIV SINGAPORE NAT.

XX Porter AG, Thanabalu T;

XX WPI: 1995-215263/28.

XX N-PSDB: AAQ87268.

XX A 4.1 bp Sau3AI-PstI sequence (given in AA087268) of B. sphaericus

CC SSI-1 included 2 ORFs, the first encoding the protein given in

CC AAR75409, and the second encoding mtx mosquitoicidal toxin (AAR75410).

XX Sequence 340 AA;

Query Match 9.6%; Score 635; DB 16; Length 340;

Best Local Similarity 39.9%; Pred. No. 5.2e-46;

Matches 130; Conservative 77; Mismatches 115; Indels 4; Gaps 1;

XX 536 ERLS---TQDEPFLQGPPIYLTLASMGISTAPYDGTASLKFALTRAKCOGKN 591

DB 6 ekmlsslrdsfvqpfllngqiltfvqmslgiasypidgytgeellkhadlanykakel9gn 65

XX 592 TYOFYRPQDSAPMLDRLTLSDLRQALTNQEFVLYFQPOVALDYGKLGVEALVRWQHPR 651

DB 66 nhrffdekmmelavlkdqiermitlalernefsvhyqdiaetgkiryfealvrvkspe 125

XX 652 LGQVAPVFIPLAEELGILNHLGQVLETCATQHOFRETRGRRLMAVNISARQFODEK 711

DB 126 lglvspeffipiaektgiltqldewwmyqaciknvelqhgfygpfllmsvnlsalqlgrad 185

XX 712 WLNSVLECLKRTGMPPELEITESLMEEDIKGTIVLLHLRREGVOVAIDDEGTGYSS 771

DB 186 fvdvkvqllnctkmkpenheleiteslivesfessicllrklngvklagddfgtgyss 245

XX 772 LSIKOLPIHRLKIDKSVNVLNMGATATIOYVIDLANGLNETVAGEISEAQOLRL 831

DB 246 lnytlilpibhklkdrslqgmtsataektliesihlahkngdvaaegvetekeyyll 305

XX 832 QKMGCHLGQGYFLTRPLPAEAMMTYL 857

DB 306 kemncdfvggyfrrpsvssdlvecl 331

## RESULT 9











PT Synechocystis species) and phycoerythrobilin conjugate, useful as  
PT fluorescent markers for biological research.

XX  
XX  
PS Example 1; Page 46; 52pp; English.

XX  
CC The present invention is related to fluorescent apophytochrome-bilin  
CC conjugates, known as phytofluors. An apoprotein known as Cph2 from  
CC Synechocystis species is used as the apophytochrome and the bilin is  
CC preferably phycoerythrobilin. The phytofluors are useful as fluorescent  
CC markers for biological research. The phytofluors have a long wavelength  
CC absorption maxima, a high molar absorption coefficient and the  
CC recombinant apoproteins can spontaneously assemble with a variety of  
CC bilin chromophore precursors. The present sequence is a phytochrome  
CC related protein from Synechocystis sp.

XX  
SQ Sequence 481 AA;

Query Match 5.3%; Score 349; DB 21; Length 481;

Best Local Similarity 32.8%; Pred. No. 5.6e-21;

Matches 80; Conservative 51; Mismatches 83; Indels 30; Gaps 4;

OY 903 WTEKIDHYVLKLERLQQRNMYEKLVLKIRASLINDIYSTVTEVROFLNWDRLVYL 962

Db 22 whrqngerrrieisikqqtgrerfingltqhirgslnetvnttaevktllqydrvli 81

OY 963 FKENSQMSGOVVTESHNDFCRSIINDEIDPCFKGHVLRLYREGSRVRAVSDIEKADLAD 1022

Db 82 yrlwgdgtgsaitesvanypsllgrtfsdevfpveyhgytkgvrainidgddleic 141

OY 1023 HKELLRHVYKANLVVY-----VFENF-----INGLTAHECKTPRYWOEDLOLLM 1070

Db 142 ladtvkqfgyvksklvplqhnrasidnesefylwqllthqcafrpwpwevelmk 201

OY 1071 ELATQVAIAIHQGLYEQLEETANIRLQOISSLDALTQGNRYLFDSTLEREQRLQRIRE 1130

Db 202 qlanqvataiqselyeqqlnkde-----nr-----vekrtdqlaatng 243

OY 1131 PLAL 1134

Db 244 slrm 247

Search completed: June 7, 2002, 18:55:13  
Job time: 267 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2002, 18:52:41 ; Search time 25.57 Seconds

(Without alignments)  
1218.892 Million cell updates/sec

Title: US-09-272-809-2

Perfect score: 6614  
Sequence: 1 MNPNSLEDPLRNVIKFKR.....GRNQWLAYEGSQLPHWDGEV 1276

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCtUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	680	10.3	752	1	US-08-309-512-8
2	680	10.3	752	5	PCT-US92-08756A-8
3	662	10.0	740	1	US-08-309-512-10
4	662	10.0	740	5	PCT-US92-08756A-10
5	651.5	9.9	765	1	US-08-309-512-5
6	651.5	9.9	765	5	PCT-US92-08756A-5
7	450.5	6.8	574	1	US-08-309-512-9
8	450.5	6.8	574	5	PCT-US92-08756A-9
9	425	6.4	580	1	US-08-309-512-6
10	425	6.4	580	5	PCT-US92-08756A-6
11	339.5	5.1	493	1	US-08-309-512-11
12	339.5	5.1	493	5	PCT-US92-08756A-11
13	299.5	4.5	499	4	US-09-090-793-11
14	297.5	4.5	525	1	US-08-375-709-19
15	297.5	4.5	525	5	US-08-752-929-19
16	257.5	3.9	748	3	US-08-904-871-6
17	257.5	3.9	748	3	US-08-904-871-13
18	245	3.7	528	3	US-08-904-871-5
19	244.5	3.7	1142	3	US-08-904-871-12
20	225	3.4	554	3	US-08-904-871-1
21	222	3.4	611	3	US-08-904-871-4
22	216	3.3	1129	3	US-08-904-871-11
23	210.5	3.2	199	5	PCT-US94-02889-2
24	207	3.1	600	3	US-08-904-871-2
25	189	2.9	600	3	US-08-904-871-3
26	143	2.2	2154	2	US-08-841-349-4
27	133	2.0	1081	2	US-08-843-530B-18

28	133	2.0	1117	2	US-08-843-530B-33	Sequence 33, Appl
29	129	2.0	2860	2	US-08-826-267-2	Sequence 2, Appl
30	127.5	1.9	15281	2	US-08-471-119A-2	Sequence 2, Appl
31	125.5	1.9	656	3	US-08-738-000-4	Sequence 4, Appl
32	125.5	1.9	656	3	US-09-258-928-4	Sequence 4, Appl
33	125.5	1.9	660	3	US-08-738-000-2	Sequence 2, Appl
34	125.5	1.9	660	4	US-09-258-928-2	Sequence 2, Appl
35	125	1.9	3248	1	US-08-353-700-1	Sequence 1, Appl
36	125	1.9	3248	5	PCT-US95-16216-1	Sequence 1, Appl
37	122	1.8	1388	4	US-09-572-191-2	Sequence 2, Appl
38	121	1.8	2482	1	US-08-328-254-6	Sequence 6, Appl
39	118.5	1.8	870	4	US-09-172-339-8	Sequence 8, Appl
40	117.5	1.8	1362	2	US-08-874-678-33	Sequence 33, Appl
41	117.5	1.8	1362	3	US-08-643-839-33	Sequence 33, Appl
42	117	1.8	655	1	US-07-736-178C-2	Sequence 2, Appl
43	116	1.8	759	4	US-09-199-637A-170	Sequence 170, App
44	116	1.8	1353	3	US-08-894-173-2	Sequence 2, Appl
45	116	1.8	1353	4	US-09-398-193-2	Sequence 2, Appl

#### ALIGNMENTS

RESULT 1  
US-08-309-512-8  
; Sequence 8, Application US/08309512  
; Patent No. 5759828  
; GENERAL INFORMATION:  
; APPLICANT: Tal, Ronny  
; APPLICANT: Benziman, Moshe  
; APPLICANT: Gelfand, David H.  
; APPLICANT: Ben-Bassat, Arle  
; APPLICANT: Calhoun, Roger D.  
; APPLICANT: Wong, Hung C.  
; TITLE OF INVENTION: CYCLIC DIGUANYLATE METABOLIC ENZYMES  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Penile & Edmonds  
; STREET: 2730 Sand Hill Road  
; CITY: Menlo Park  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 94025  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/309,512  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/800,218  
; FILING DATE: 29-NOV-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bortner, Scott R.  
; REGISTRATION NUMBER: 34,298  
; REFERENCE/DOCKET NUMBER: 8145-008  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 854-3660  
; TELEFAX: (415) 854-3694  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 752 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEetical: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Acetobacter xylinum

US-08-309-512-8

Query Match 10.38; Score 680; DB 1; Length 752;  
Best Local Similarity 26.28; Pred. No. 9e-55;  
Matches 228; Conservative 130; Mismatches 323; Indels 188; Gaps 20;

```

0Y 19 HRLTFRLETOIYEARIFLGVDRVITYKFASSGESEVIAEAVNRALS-----L 70
Db 11 HGAVGRADVLEFYTDOKRI-----SADVLLSMEQAIIDPMVYIDENHLL 54
0Y 71 LGLHPVEDIPPOAREBELGNOKKMAIVDAVHRRKSHLS-----GRISPT-----E 117
Db 55 IFFNAAEKIMGCSREYEMGHVNSCLPEPERDORHDDYIRNRNETGVGRIVGTSREVERA 114
0Y 118 HSNCHY-----TIVDSCHIOYLLAMGVLSLTVPVNODQOLGIMAVHHSKRFRF 168
Db 115 RANBEYVCGELSTIRVOYNDGKIYYTA-----VKMD-----VTESROKRII 157
0Y 169 EOEETMALLSKEVSLAITOSLSROYHOOOVEALVORLETTYVAOYGDREPMOYALET 228
Db 158 VLONDVLOALASDMSI-----DEVGDLICRADSFVP----- 189
0Y 229 VGOAEVADGANLYIARPDLTGSVAOHTYOMNLRFPDMGNMLETSLMOELRGQPSAMERMA 288
Db 190 -----GAAGALMLT-----DSSRLSVS-----ASPSMKRYRAA 219
0Y 289 VOSTWEMKRPPTSYA--PLPPTNCVPHPGTYLGELEQSRDMIAPPELSAENFOSFLIPL 346
Db 220 LDSMQLPPEQJEMLRANPAGSNVYMPDGYA-----SLARSJGLERCCSSPTISR 268
0Y 347 AADQOWGSLILLKREKSLYKHNAGKKGIDRRNILPLSPAEAMEFYOKLYPTNRSERKL 406
Db 269 TGOVWGFALYLRGDEDEL-----AMAO----- 291
0Y 407 AQVASTOL-----YMAITOFOYFRLTQOTAAVDPLTOLEPNMIIENROLTALDALLY--EG 460
Db 292 -RIVYSTMPFCALAIEOSETPHOIAOLSNDLSLGL-----LNRSLHNIEHLINRGGD 345
0Y 461 KMGVLYIAMDREKRINESFGHKTGDQLLOVDADRLNOKLSPLAANSPLLSRNHGDSFTI 520
Db 346 SQFSLFVWDIDRFNDINADLGHVADRFLEIGRRH-----LVKDDYIYRSRGGGEFII 401
0Y 521 LLTOSIDMOEMIPLCERNLSTFFQEPFLQOGPYLTATSMGISTAPYDGEAESLTPAEI 580
Db 402 VVPECS--HERAEKAEUNLINAIPARLOYGENTLSTISCCVISTFFPANGPSELSHAD 460
0Y 581 ALTRKACOGKNTYOFYRPRODASPMIDRLTESDLRQALNTQOEVLVYFOVALADTSGLL 640
Db 461 ATROAKEDGRGLFPRFAGOEKQOVADRLVIGSALRDSLSTGMLNLTNYQOVETMTGGLVG 520
0Y 641 VEALVRQNHPRYGOVARDVFTPLABEGLINHLGQWVLETAACATHQHFPEFETGRRLMAY 700
Db 521 VEALSRMNHPRPLGNIYPSRFJAAVEAGLEIGRMSLEBSCQOMVAKMDGDGVAPVPA 580
0Y 701 NISAROFODEKMLSVLECLKRTGMPREDLELETESIMEDIKGTVLLHRLREGEVOY 760
Db 581 NLSAVHPRNRCLPREHIANLKLHNRGLDRDLVETTESVMDSSSETEVUHAIRRLDYL 640
0Y 761 AIDDEGTGYSLSILKOLPIHRLKIDKSFYVNDLNEGADTAIIQYVIDLANGLETVAE 820
Db 641 SMDPEGTGYSLSLRLTRPLTEIKIDSSFINDEPHDINAQVMAVYIGISRLGMVTE 700
0Y 821 GISEBAQOLRQKKMGCHLGOGYFLTRPLP 849
Db 701 GVEIEQOWRLLEELHCDVMOGYFLSKRLP 729

```

APPLICANT: Benzman, Moshe  
 APPLICANT: Gelfand, David H.  
 APPLICANT: Ben-Bassat, Arie  
 APPLICANT: Calhoun, Roger D.  
 APPLICANT: Wong, Hing C.  
 TITLE OF INVENTION: CYCLIC DIGUANYLATE  
 TITLE OF INVENTION: METABOLIC ENZYMES  
 NUMBER OF SEQUENCES: 11  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Limbach and Limbach  
 STREET: 2001 Ferry Building  
 CITY: San Francisco  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 94111  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0,  
 SOFTWARE: Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US92/08756A  
 FILING DATE: 19921014  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/800,218  
 FILING DATE: 29-NOV-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Borthner, Scott R.  
 REGISTRATION NUMBER: 34,298  
 REFERENCE/DOCKET NUMBER: WER 20030 USA  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-433-4150  
 TELEFAX: 415-433-8716  
 TELEX: 278356  
 INFORMATION FOR SEQ ID NO: 8:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 752 amino acids  
 TYPE: AMINO ACID  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHEICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Acetobacter xylinum  
 PCT-US92-08756A-8

Query Match	10.3%	Score 680	DB 5	Length 752
Best Local Similarly	26.2%	Pred. No. 9e-55		
Matches 228	Conservative 130	Mismatches 323	Indels 188	Gaps 20

```

OY 19 HRALTRFETLOVYIEENAFIFLGVDVRKIKYKASDGSSEVLAENNRALRS-----L 70
Db 11 HGAGRAVLEFVYDQPT-----SADVLISAEQALDPVVIDENHLI 54
OY 71 LGIHEPVEDIPOAREELIGNOKMIADVAAHRRKSHLS-----GRISPT-----E 117
Db 55 IFFNNAEKINGCSHEEYMGNNVSCVPEPERDRHDDYINNRETVGIRIYTSREVEFR 114
OY 118 HSNHGY-----TYYDSCHIDYLLAMGVSSLYPVWQDOOLKIMAVHNSKPRFT 168
Db 115 RANGFYVCGELISIRVOVNDGKIYYIA-----VMKD-----VTEOSROKIL 157
OY 169 EOEWEFTMALISKEVSLATIOSQLSRQVHOQOVCEALYORLETTYAQYGDREPTWQYALET 228
Db 158 VLOQNVLOALASDMSI-----QEVGDLLOCRADSFVP----- 189
OY 229 VGOAVEADGAVLYIAPDLTGSVAOHTYQWNLRFDMGNNLETSLMOELMRGQPSAAMEPMA 268
Db 190 -----GAVGALMLI-----DSRRISYS-----ASPSMKRYRAA 219

```

QY 289 VOSTWEKPRPTSYA--PLPPTNCVPHGYTIGELBQRSDWIAAPPSLSAENFQSLIYPL 346  
Db 220 LDSQAPPEQLEMLRANAGSNVYMDYA-----SLARSLGLERCCSSTIIR 268  
QY 347 AADQWWSLLILRKREKSLVHMAGKRGIDRRNIIPLRSFEAMETQKLVPTWNSERKL 406  
Db 269 TGVWNGVAFALYLRGDEAL-----AMAO----- 291  
QY 407 AQAASDOL--YMAITQOFTRLITQOTAYDPLTQLPNMIIFNRQTLALLDALY--EG 460  
Db 292 -RVYSTSMPCALAIQESQTHIAQLSNFDSLGL-----LNRSLNHTIERLIMRGD 345  
QY 461 KMGVLYIAMRFRKINESFGHKTGDLQEVADRNLKSLPLAAYSPLSRMHGDGTI 520  
Db 346 SQSLFMYDIDFRDINDALGHVADRFLIIGRIRH---LVKDDYIVSRSGDEPII 401  
QY 521 LITQISDQEMIPICERLSTFQEPFLOGOPILYLTASMGISTAPYDETFESLTKFAEI 580  
Db 402 VVPECS-HERAKRAENILNIAARPLQVAGENTISCCVGISTFPANPDSESLSHADA 460  
QY 581 ALTRACQGNKYQFYRPQDSAPMLDLTLLESDLRQALTNQEFVLYFQPOVALDTGKLLG 640  
Db 461 ATROAKEDGRGLFRPAGQEKQNOADRLVLSALRDSLKGLMLNLYQOYETMTGGLYG 520  
QY 641 VEALVRQHPRLGOVAPRVFTPLABELGLINHLGQWLETACATHQHFRETGRRLMAY 700  
Db 521 VEALSRMHHPRLGNITPFRFAVAEETGOIEARMSLEACQWVKVDRGVPRPTAV 580  
QY 701 NISAROFODEKMLSVLECKRTGCMRPEDLELETESLMMEDIKGTVLLHRLREGVQV 760  
Db 581 NLSAVHRNRRLPREHIANLKHNGITPRRLVEITTESYMMSSSTTEVYLAIIRLDYGL 640  
QY 761 AIDDFGTGYSSLILKOLPIHRLKIDKSFVNDLNEGADTAIIOYVIDLANGLMEYAE 820  
Db 641 SMDFGTGYSSLSLRTPLRLEIKIDRSFINDFEHDNTAQAIVTMAYIGISGLMTVYTE 700  
QY 821 GIESEAOLOKXMGCHGOGYFLTRPLP 849  
Db 701 CVTEQOQRLLELHCDVMQGLFPSKPLP 729

RESULT - 3  
US-08-309-512-10  
; Sequence 10. Application US/08309512  
; Patent No. 5759828  
; GENERAL INFORMATION:  
; APPLICANT: Tai, Ronny  
; APPLICANT: Benziman, Moshe  
; APPLICANT: Gelfand, David H.  
; APPLICANT: Ben-Bassat, Arie  
; APPLICANT: Calhoun, Roger D.  
; APPLICANT: Wong, Hing C.  
; TITLE OF INVENTION: CYCLIC DIGUANYLATE METABOLIC ENZYMES  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 2730 Sand Hill Road  
; CITY: Menlo Park  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 94025  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/309,512  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/800,218  
; FILING DATE: 29-NOV-1991

ATTORNEY/AGENT INFORMATION:  
NAME: Bortner, Scott R.  
REGISTRATION NUMBER: 34,298  
REFERENCE/DOCKET NUMBER: 8145-008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 854-3660  
TELEFAX: (415) 854-3694  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 740 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEetical: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Acetobacter xylinum  
US-08-309-512-10

Query Match 10.0%; Score 662; DB 1; Length 740;  
Best Local Similarity 26.3%; Pred. No. 4.3e-53;  
Matches 207; Conservative 143; Mismatches 296; Indels 140; Gaps 20;

QY 115 PTEHNGHYTTVD-----SCHIOYLAMGYL---SSLTVYVQDQOLM--G 155  
Db 59 PTLHOAGHDAVERSGSSHNRIYVGTSHREVEFTREDGEYICGLSLSVYVNDKRIFFMG 118  
QY 156 IM-AVHHSKRPRFQEWEMTALLSKREVSALITQSLSROYHQOQVQDALYQRETTVA 213  
Db 119 VMKNVTNESQQRK-----LILQNDVLAQALASDMAI---QDVADLCRRVSEFV- 164  
QY 214 QYCDRPFMOYALETQCAVEADGAVLYIADPLTGSVAQHOMNLFPMGMWLETSLMOE 273  
Db 165 -PGT-----YAVMLITPPDQGLRVLS--SPTLKRRAISLE---SLYSSSELEK 208  
QY 274 LMRGQPSAAMEPMAVOSTWEKRPPTSVAPLPPTNCVPHGYTIGELBQRSDWIAAPPSL 333  
Db 209 L-----RVDPKHATRVWDSYRSL-----GISLG----- 232  
QY 334 SAENFQSLIYPLAADQWWSLLILRKREKSLVKHMAGKRGIDRRNIIPLRSFEAMETQ 393  
Db 233 -LQCCFCIPVSTRSQVKGIFALYSRDEQGRTOFORIVDSCIPPCALAFEO----- 283  
QY 394 KLVPTWNSERKLAQVASTQ-----LYMAITQOFTRLITQOTAYDPLTQLPNMIIFNRQ 448  
Db 284 -----NATQEHISLANFDLITGLNRSVHKVIEGMSIQDG-----NRQ 324  
QY 449 LTLALLDALYEGKMGVLYIAMDRKRNESFGHKTGDLQEVADRNLKSLPLAAYSP 508  
Db 325 FAIFMLD-----IDRPDINDALGHVADQFLIEIARIRS-----IAKBDY 366  
QY 509 LLSRNHGGFTLITQISDQEMIPICERLSTFQEPFLOGOPILYLTASMGISTAPYDE 568  
Db 367 VLSRSGDEFFVYVVD-CPRKEATDFAEHLASMTMQIQNTLTLSIGISTPYPDG 425  
QY 569 ETASLTKFAEIALTRACQGNKYQFYRPQDSAPMLDLTLLESDLRQALTNQEFVLYFQ 628  
Db 426 PDSESLSTADVALRQAKEDGRGVFRANLEKNOVADRLVLSALRDSLKGLMLNLYQOYETMTGGLYG 485  
QY 629 POVALDTGKLLGEALVRQHPRLGOVAPRVFTPLABELGLINHLGQWLETACATHQHF 688  
Db 486 FOVRTHTLELSGVEALSWMHHPRLGNITPFRFAVAEETGOIEARMSLEACRQIYKW 545  
QY 689 FRETGRRLMAYVNTSAROFODEKMLSVLECKRTGCMRPEDLELETESLMMEDIKGTIV 748  
Db 546 DDDGIHVPTAVVNLNSAVHRNRRLPREHIANLKHNGITPRRLVEITTESYMMSSSTTEVYLAIIRLDYGL 605  
QY 749 LILHRLREGVQVADIDFGTGYSSLILKOLPIHRLKIDKSFVNDLNEGADTAIIOYVID 808  
Db 606 VLQSTRNIGCGLSMDDFGTGYSSLSRLRLPLRLEIKIDRSFINDFEHDNTAQAIVTMAYIG 665

QY 809 LANGINLETVAEGIESEAOLOKMGCHLGOGYFLTRPLPAEAMTYLYYPQ-ILDFGP 867  
Db 666 IGSRIGMTVTEGVEEQDRLKNCVMOGYLFAKPLAPDDEKWMRHQTIROMLP 725  
QY 868 TPLPK 873  
Db 726 AAPAK 731

## RESULT 4

PCT-US92-08756A-10  
Sequence 10, Application PC/TUS9208756A  
GENERAL INFORMATION:  
APPLICANT: Tal, Ronny  
APPLICANT: Benziman, Moshe  
APPLICANT: Gelfand, David H.  
APPLICANT: Ben-Bassat, Arie  
APPLICANT: Calhoon, Roger D.  
APPLICANT: Wong, Hing C.  
TITLE OF INVENTION: CYCLIC DIGUANYLATE  
TITLE OF INVENTION: METABOLIC ENZYMES  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Limbach and Limbach  
STREET: 2001 Ferry Building  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0,  
SOFTWARE: Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/08756A  
FILING DATE: 19921014  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/800,218  
FILING DATE: 29-NOV-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Bortner, Scott R.  
REGISTRATION NUMBER: 34,298  
REFERENCE/DOCKET NUMBER: WEYR 20050 USA  
TELEPHONE: 415-433-4150  
TELEFAX: 415-433-8716  
TELEX: 278356  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 740 amino acids  
TYPE: AMINO ACID  
MOLECULE TYPE: linear  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Acetobacter xylinum  
PCT-US92-08756A-10

Query Match 10.0%; Score 662; DB 5; Length 740;

Best Local Similarity 26.3%; Pred. No. 4,3e-53; Indels 140; Gaps 20;

Matches 207; Conservative 143; Mismatches 236; Indels 140; Gaps 20;

QY 115 PTEHNGHYTTVD-----SCHIQYLLAMGYL---SLTVPVNQDQOLW--G 155  
Db 59 PTLHAGHDADVERSSSHNRIVGTSREVEFTRSDDGEYIGCELSTKSVVNDKRIFFMG 118  
QY 156 IM-AVHHSKRRTTEDEWEMTALLSKEYSLATIQSOLSRQVHOQOQVDEALVORLETTVA 213

Db 119 VMKNVTNESQORK-----ILLQNDVLOALASDMKI-----QDVADLLCRVESFV- 164  
QY 214 QYDREPETWOYALETVGOAVADGAVLYIAPDLTGSVQHYQWNLRFPMGNMLETSLMOE 273  
Db 165 -----FQT-----VAVMLILTPDGLRVLS---SPTLKRYSLE---SLYSSSELEK 208  
QY 274 LMRGOPSAMEBMAVOSTWEKPRPETSVAPLPTNCVPHGYTLGELRQSDWIAPPESL 333  
Db 209 L-----RVPKHKATRWVWDSYRSL-----GISLG----- 232  
QY 334 SAENFOSFLIYPIADQQWGSLLILRKEKSLVKKHAKGRIGDRNNILPRISFEAMERTQ 393  
Db 233 -----LQCCFCTPVSTRSGQGITRALSREDQGRTOPORIVDSICIPCALAFQ----- 283  
QY 394 KLVPWNRSERKLAOVASTQ-----LYMAITQOEVYTRLIQOTVADPLTQLPNNITFNRQ 448  
Db 284 -----NATQEHISHLANFDSLTLGLNLSVHKYIEGMSIQDG-----NRQ 324  
QY 449 LTLALLDLYEGKMGVLIAMDRKRNESFGKGTGGLLOEVADRINOKLSPLAAYSP 508  
Db 325 FAIFMLD-----IDREINDALGHVYADQFLIEIARIRS---IAKEDY 366  
QY 509 LLSRWGSGFTILLQISDNQEMIPICERLLSTQEPFLQGPYLYLASMISTAPYD 568  
Db 367 VLSRSGDEFFVWVPD-CPHKEATDFAEHLASMTMPQIQNTLTTSIGISTYPPDNG 425  
QY 569 ETASLTKFAEIALTRAKCGKNTYQFYRPQDSAPMLDRLTLESPLROALTNQEFVLYFQ 628  
Db 426 PDSESLSTADVALROAKEDRGVFRFANLEKNVOAQRVLGSLRSLAKGMNLHLYQ 485  
QY 629 POVALDTGKLIGVEALVWQHPRLGQVADVFIPLAEFLGLNLHGMVLETACATHQF 688  
Db 486 PQVTHLTLELGSVALSRHHPHGLNIPFSRFIAVEETGOIEAIGRSLLEACRQIYKW 545  
QY 689 FRETGRIRMAVNTSARQODEKMLNSVLECKTKFGMPDELTLETPSLMEDIKGVV 748  
Db 546 DRDGIHVPFTAVNLSAVHRRNALPETHAALLKDHNLPSLTYLETIESVWMDSRDTEE 605  
QY 749 LILRLREGVOVAIDFEGTGYSSILKQDPIHRLKIDKSFVNDLNGADTAIIOVYID 808  
Db 606 VLOSIRNIGCGLSMDPDFGTGYSLSRLRLPLTEKIDRSFINDEHDTNAQAVTAMVIG 665  
QY 809 LANGINLETVAEGIESEAOLOKMGCHLGOGYFLTRPLPAEAMTYLYYPQ-ILDFGP 867  
Db 666 IGSRIGMTVTEGVEEQDRLKNCVMOGYLFAKPLAPDDEKWMRHQTIROMLP 725  
QY 868 TPLPK 873  
Db 726 AAPAK 731

## RESULT 5

US-08-309-512-5  
Sequence 5, Application US/08309512  
Patent No. 5759628

GENERAL INFORMATION:

APPLICANT: Tal, Ronny  
APPLICANT: Benziman, Moshe  
APPLICANT: Gelfand, David H.  
APPLICANT: Ben-Bassat, Arie  
APPLICANT: Calhoon, Roger D.  
APPLICANT: Wong, Hing C.  
TITLE OF INVENTION: CYCLIC DIGUANYLATE METABOLIC ENZYMES  
NUMBER OF SEQUENCES: 63  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 2730 Sand Hill Road  
CITY: Menlo Park  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 94025

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk



[illegible]



APPLICATION NUMBER: US 07/800,218  
FILING DATE: 29-NOV-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Bortner, Scott R.  
REGISTRATION NUMBER: 34,298  
REFERENCE/DOCKET NUMBER: WEYR 20050 USA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-433-4150  
TELEFAX: 415-433-8716  
TELEX: 278356  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 574 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Acetobacter xylinum  
PCT-US92-08736A-9

Query Match 6.8%; Score 450.5; DB 5; Length 574;  
Best Local Similarity 25.4%; Pred. No. 2.3e-33;

Matches 136; Conservative 111; Mismatches 208; Indels 81; Gaps 14;

QY 346 LAADQWVSGLLLRKESLVKHNAGKR--GIDRRNLLPRLSFEAMEETOKLYPTWNRS 402  
DB 82 IKAQGLMGNICNRAKQDSL--YVWATTIIPKIDROGITTYVASRFEETELM-----NT 134  
QY 403 ERKLAQVASTOLYMAITQOFTRLITQOTATDPLTQLEPNWIIENQTLALDALIEGKM 462  
DB 135 RDRICELET-----DPLTGLNRGGFN--TALADELTARCRE 169  
QY 463 VG-----VLVIADRFKRNESFGKTDGGLLOEVADRNLNOKSLPLAAYSPILSRMGD 516  
DB 170 PGMTHAPRLAFDDGFGQINDVGHNGADIVLRAIASRLIEHTP---DDPVSRLEGD 225  
QY 517 GTTILLQISDQEMIPLCERILSTPOEPFLQGGPIYLTASMGISTAPYDG-ETAESIL 575  
DB 226 EFAYILHRTLEDVSLERMDRLQALIERIDIEVTYVSAGSIG--AVLIDGTDTIMEVQ 283  
QY 576 KFAIETLRACQCKNTYQFRPODSAPMLDRILLESPLRQALITQOEVLVYFQPOVALDT 635  
DB 284 KNAWMAWMAARAGKQSQMFTRLRERARVSTLSEARCGVENQPEVYQPIVNCNT 343  
QY 636 GKLLGVETLVWQHPRLGQVAPDVFIPLAEELGLINHLGQWLET---ACATHQHFRE 691  
DB 344 MEVDQIEALLWQHPERGLLAEDSDVFTDAGLAQAGPRRIEAFRRDVC---MWNE 398  
QY 692 TGRRLR-MAVNISAROFODEKWLNSVLECLKRTGMPPEDELEITESIMMEDIKGTVAL 750  
DB 399 KGGPRQALINLSMDLIRDDYQRELESRLRFPNSPDSVLEVTEA-----ML 447  
QY 751 HRLAEE-----GYOVAIDEGTGYSSILKQDPIHRLIKDKSFVNDLNEGA 798  
DB 448 HGRREEGIRMLRELARGFRIALDNFGKITVNLHIRELFPQVKIDQSWMTNIVGMD 507  
QY 799 DTAITQVYIDLANGLNETVAEGTSEAOQORLOKMGCHLQOGYTLTPRLPAEAM 854  
DB 508 ACWVLSLIDMGQGFNMVEVTEGVENRQDFLVALRPERIOGFVSSALSSODTL 563

# RESULT 9

US-08-309-512-6  
Sequence 6, Application US/08309512  
Patent No. 5739828  
GENERAL INFORMATION:  
APPLICANT: Tal, Ronny  
APPLICANT: Ben-Zaman, Moshe  
APPLICANT: Gelfand, David H.  
APPLICANT: Ben-Bassat, Arie

APPLICANT: Calhoun, Roger D.  
APPLICANT: Wong, Hing C.  
TITLE OF INVENTION: CYCLIC DIGUANYLATE METABOLIC ENZYMES  
NUMBER OF SEQUENCES: 63  
CORRESPONDENCE ADDRESS:  
ADDRESS: Pennie & Edmonds  
STREET: 2730 Sand Hill Road  
CITY: Menlo Park  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 94025  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/309,512  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/800,218  
FILING DATE: 29-NOV-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Bortner, Scott R.  
REGISTRATION NUMBER: 34,298  
REFERENCE/DOCKET NUMBER: 8145-008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 854-3660  
TELEFAX: (415) 854-3694  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 580 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Acetobacter xylinum  
US-08-309-512-6

Query Match 6.4%; Score 425; DB 1; Length 580;  
Best Local Similarity 26.1%; Pred. No. 6e-31;

Matches 133; Conservative 102; Mismatches 226; Indels 48; Gaps 10;

QY 352 WVGSLILRKEKSLVKNAGKRGIDRRNLLPRLSFEAMEETOKLYPTWNRSEKLAQVAS 411  
DB 88 WGNICNRAKQDGL--YVWATTIMPKHNSLG--AVEGYVATFEETELMNTDRILKSLDA 143  
QY 412 TQLVMAITQOFTRLITQOTATDPLTQLEPNWIIENQTLALDALIE-CKWVGVLVAM 470  
DB 144 T-----DPLTGLNRGGFNVLQTAVEDKSQNTIRIMLVMDL 182  
QY 471 DFRKRNESFGKTDGGLLOEVADRNLNOKSLPLAAYSPILSRMGDGTIILLQISDQOE 530  
DB 183 DFGKQINDVGHNGADIVLRAIASRLIEHTP---DDPVSRLEGD 225  
QY 531 MIPLCERILSTPOEPFLQGGPIYLTASMGISTAPYDG-ETAESILKFAIETLRACQCK 590  
DB 239 LSLMLEKLLAELEAVIEGNTVNVNVSIGV--TPIASQSAESILQKNADIALYAKRAGG 297  
QY 591 NTYQFRPODSAPMLDRILLESPLRQALITQOEVLVYFQPOVALDTGKLLGVETLVWQHP 650  
DB 298 HQARFEDMTLHQHALERQQLINDAREGVKQDFELYTOPINNFSTGKCDQLEALLPWHBP 357  
QY 651 RLQGVAP-----DVFIPLAEELGLINHLGQWLETACATHQHFRETGRRL---MAVNI 702  
DB 358 QRLGLAASFRDVLIDAA-----LAQVMSPRLYKSFQNDMRMWNSTLDAIPLTLTINL 409  
QY 703 SAROFODEKWLNSVLECLKRTGMPPEDELEITESLMEDIKGTIV-LIHLRREGQVYA 761





GENERAL INFORMATION:  
APPLICANT: Calgene, LLC  
TITLE OF INVENTION: Production of polyunsaturated fatty acids by expression  
TITLE OF INVENTION: of polypeptide-like synthesis genes in plants  
FILE REFERENCE: CGNE.131.01US  
CURRENT APPLICATION NUMBER: US/09/090,793  
CURRENT FILING DATE: 1998-06-04  
EARLIER APPLICATION NUMBER: 60/048,650  
EARLIER FILING DATE: 1997-06-04  
NUMBER OF SEQ ID NOS: 66  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 11  
LENGTH: 499  
TYPE: PRT  
ORGANISM: Shewanella putrefaciens  
US-09-090-793-11

Query Match 4.5%; Score 299.5; DB 4; Length 499;  
Best Local Similarity 29.0%; Pred. No. 3.1e-19;  
Matches 101; Conservative 51; Mismatches 129; Indels 67; Gaps 11;

QY 933 KIRASLINDILYSTVTEVROFLNTD-----RVVLFKFNQSGGVVTESHNDFCRSIIN 987  
DB 182 RIGANTKLNKY---TATSERLITNVGKPLKLVYHTNNOPPMI-----DYSIIILL 232

QY 988 DEIDPCFEGHYLRLY---REGGRVAVSDIEKADLADCHKEILRYHYOKANLVVPEVNE 1044  
DB 233 VEMSLFLILAYFLYSFLVLPYR-KLASDIKMDKSREIKRLRYHYPITELKVAHTN- 290

QY 1045 NIMGLLIAHECKTPRYWOEDLQIMELATQVAIAIHOGELYEODETANIRLQOISSIDA 1104  
DB 291 -----ALM-----GTIOEQTKOLN---EGV-FIDK 311

QY 1105 LTQVGNRLFLDSTLEREMORLORIREPLALLICDVDFKGFNDNGHPAGDRCLKKIDA 1164  
DB 312 LTNIENRRAFERLETYQOLARQOIGFTLLIADVDHREKYNDTGLHAGDALILKVAOT 371

QY 1165 MAKVAKRPDILVARYGGEFATIISETSLGAINTVEALQVEVANIAPHTVSGT-GHVT 1223  
DB 372 LSOQFYRAEDICARFGGEFFIMLFRDIPDEPLQKLDMLHSFALNLPHPNSTANVT 431

QY 1224 LSGIAV-----YTPERHINPNALVKADLALYEAKAGRNQMLA 1263  
DB 432 VSLGVCYVAVDDEFEKSESHIGSOALLADKALYHAKACGRNALS 479

RESULT 14  
US-08-375-709-19  
Sequence 19, Application US/08375709  
Patent No. 5683898  
GENERAL INFORMATION:  
APPLICANT: YAZAWA, Kazunaga  
APPLICANT: YAMADA, Akiko  
APPLICANT: KATO, Seishi  
TITLE OF INVENTION: Gene Coding for Elcosapentaenoic Acid  
TITLE OF INVENTION: Synthesizing Enzymes and Process for Production of  
TITLE OF INVENTION: Elcosapentaenoic Acid  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/375,709  
FILING DATE: 20-JAN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/178,251  
FILING DATE: 14-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-147945  
FILING DATE: 15-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: WEGNER, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 53466/150/AAOK  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 525 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-375-709-19

Query Match 4.5%; Score 297.5; DB 1; Length 525;  
Best Local Similarity 29.0%; Pred. No. 5.2e-19;  
Matches 100; Conservative 50; Mismatches 128; Indels 67; Gaps 11;

QY 933 KIRASLINDILYSTVTEVROFLNTD-----RVVLFKFNQSGGVVTESHNDFCRSIIN 987  
DB 208 RIGANTKLNKY---TATSERLITNVGKPLKLVYHTNNOPPMI-----DYSIIILL 258

QY 988 DEIDPCFEGHYLRLY---REGGRVAVSDIEKADLADCHKEILRYHYOKANLVVPEVNE 1044  
DB 259 VEMSLFLILAYFLYSFLVLPYR-KLASDIKMDKSREIKRLRYHYPITELKVAHTN- 316

QY 1045 NIMGLLIAHECKTPRYWOEDLQIMELATQVAIAIHOGELYEODETANIRLQOISSIDA 1104  
DB 317 -----ALM-----GTIOEQTKOLN---EGV-FIDK 337

QY 1105 LTQVGNRLFLDSTLEREMORLORIREPLALLICDVDFKGFNDNGHPAGDRCLKKIDA 1164  
DB 338 LTNIENRRAFERLETYQOLARQOIGFTLLIADVDHREKYNDTGLHAGDALILKVAOT 397

QY 1165 MAKVAKRPDILVARYGGEFATIISETSLGAINTVEALQVEVANIAPHTVSGT-GHVT 1223  
DB 398 LSOQFYRAEDICARFGGEFFIMLFRDIPDEPLQKLDMLHSFALNLPHPNSTANVT 457

QY 1224 LSGIAV-----YTPERHINPNALVKADLALYEAKAGRNQ 1260  
DB 458 VSLGVCYVAVDDEFEKSESHIGSOALLADKALYHAKACGRNQ 502

RESULT 15  
US-08-752-929-19  
Sequence 19, Application US/08752929  
Patent No. 5798259  
GENERAL INFORMATION:  
APPLICANT: YAZAWA, Kazunaga  
APPLICANT: YAMADA, Akiko  
APPLICANT: KATO, Seishi  
TITLE OF INVENTION: Gene Coding for Elcosapentaenoic Acid Synthesizing  
TITLE OF INVENTION: Enzymes and Process for Production of Elcosapentaenoic  
TITLE OF INVENTION: Acid  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.

COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/752,929  
FILING DATE: 20-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/375,709  
FILING DATE: 20-JAN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/178,251  
FILING DATE: 14-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-147945  
FILING DATE: 15-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: WEGNER, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 53466/150/AAOK  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 525 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-752-929-19

Query Match 4.5%; Score 297.5; DB 1; Length 525;  
Best Local Similarity 29.0%; Pred. No. 5,2e-19;  
Matches 100; Conservative 50; Mismatches 128; Indels 67; Gaps 11;

OY 933 KIRASTINDILYSTVEYRQFLNTD-----RYVLKFNFSNMSGGVVTVESHNDPCRSIIN 987  
DB 208 RIGANTKLNKV---TATSERLITNWDGKPLKLVLYHTNNOPPMVL-----DYSLIILL 258  
OY 998 DEIDDPCEKGYRLRY---REGRVAVSDIEKADLADCHKELLRHYQYKANLVVPVENE 1044  
DB 259 VEMSFLLILAYFLYSYELVPRV-RLASDIKKMKSREIKLRHYHPITELVKVATHFN- 316  
OY 1045 NIMGILLIAHECTPRYQWEDLQIMELATQVAIAIHOGELYEQLETTANIRLQOISSLDA 1104  
DB 317 -----ALM-----GTIOEOTKOLN---EOV-FIDK 337  
OY 1105 LTQVGNRYLFDSTLREMQRLQRIREPRLALLCDVDPEKGFNDNNGHPAGDRCLKKIADA 1164  
DB 338 LTNINRRAFERLRYCOLLRQOIGFTLLIADVDHEKRYNDTGLHAGDEALIVAQOT 397  
OY 1165 MAKVAKRPDLVARYGGEFFAILLETSEGAINTTEALQVEANLAIPHTVSGT-GHYT 1223  
DB 398 LSQGFYRAEDICARFGGEFFIMLFRDIPDEPLQRIKIDAMLHSAELNLPHPNSSTANYVT 457  
OY 1224 LSIGIAV-----YTPRHINPNALVKADLALYEAKAKGRNQ 1260  
DB 458 VSLGVCYVAVVDPEFKSSESHITGSOALIIDKALYHAKACGRNQ 502

Search completed: June 7, 2002, 18:55:44  
Job time: 183 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2002, 18:53:06 ; Search time 37.51 Seconds  
(without alignments)  
3268.728 Million cell updates/sec

Title: US-09-272-809-2

Perfect score: 6614  
Sequence: 1 MNPNSLEDFLRNVINKFHR.....GRNQWLAYEGSQLPHVDGEV 1276

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR.71:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6614	100.0	1276	2 S75801	probable phytochro
2	967.5	14.6	880	2 AD1953	hypothetical prote
3	922	13.9	1415	2 C83070	conserved hypotet
4	882	13.3	748	2 D87632	sensory box/GGDEF
5	865	13.1	1244	2 S76102	hypothetical prote
6	862.5	13.0	1245	2 H83574	conserved hypotet
7	856.5	12.9	565	2 AB2334	hypothetical prote
8	830.5	12.6	760	2 E83610	conserved hypotet
9	821.5	12.4	696	2 S75626	hypothetical prote
10	819.5	12.4	788	2 AF0122	probable membrane
11	813.5	12.3	783	2 B83232	hypothetical prote
12	811.5	12.3	685	2 H83428	conserved hypotet
13	811	12.3	899	2 A83019	conserved hypotet
14	810.5	12.3	687	2 F83100	conserved hypotet
15	809	12.2	840	2 S74707	nitrogen fixation
16	801.5	12.1	840	2 S76238	hypothetical prote
17	797	12.1	1578	2 AC2094	two-component resp
18	780	11.8	611	2 G82965	conserved hypotet
19	768	11.6	951	2 E82297	c-di-GMP phosphodi
20	757	11.4	865	2 E75292	GGDEF family prote
21	752.5	11.4	856	2 AH3133	GGDEF family prote
22	749.5	11.3	724	2 D98154	hypothetical prote
23	749.5	11.3	737	2 H83386	conserved hypotet
24	746	11.3	864	2 A96297	hypothetical 91.8K
25	741	11.2	644	2 AF2986	GGDEF family prote
26	739.5	11.2	732	2 S77565	hypothetical prote
27	737.5	11.2	821	2 AI2417	hypothetical prote
28	737.5	11.2	821	2 AI2417	hypothetical prote
29	736	11.1	709	2 AE3468	di-guanilate cyclas

30	731.5	11.1	842	2 A87341	sensory box/GGDEF
31	729	11.0	772	2 G98327	hypothetical 91.8K
32	729	11.0	772	2 AF2955	GGDEF family prote
33	727.5	11.0	752	2 H98161	hypothetical 91.8K
34	727.5	11.0	779	2 A83125	GGDEF family prote
35	717.5	10.8	564	2 E85895	hypothetical prote
36	714	10.8	742	2 H95270	probable kinase/es
37	714	10.8	1051	2 C95367	conserved hypotet
38	708	10.7	1021	2 AC2202	hypothetical prote
39	705.5	10.7	805	2 B75515	sensory box/GGDEF
40	703.5	10.6	749	2 A69861	conserved hypotet
41	700.5	10.6	809	2 B87260	sensory box/GGDEF
42	699.5	10.6	746	2 T35811	probable phosphodi
43	688.5	10.4	696	2 G97626	hypothetical prote
44	688.5	10.4	696	2 AI2849	GGDEF family prote
45	687	10.4	543	2 C84021	hypothetical prote

## ALIGNMENTS

RESULT 1  
S75801  
probable phytochrome sl10821 - Synechocystis sp. (strain PCC 6803)  
C:Species: Synechocystis sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
C:Accession: S75801  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys  
s.  
A:Reference number: S74322; MUID:97061201  
A:Accession: S75801  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1276 <KAN>  
A:Cross-references: EMBL:D64003; GB:AB001339; NID:g1001200; PIDN:BA10536.1; PID:d101  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C:Superfamily: phytochrome homology  
C:Keywords: chromoprotein; phytochromobilin  
F:1-416/Domain: phytochrome homology #status atypical <PHT>  
F:129/Binding site: phytochromobilin (Cys) (covalent) #status predicted

Query Match	100.0%; Score 6614; DB 2; Length 1276;	Best Local Similarity	100.0%; Pred. No. 0;	Matches 1276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	1	MNPNSLEDFLRNVINKFHRALTLRETLQVIVEARIFLGVDRVITYKSPASDGSCEVLAE	60	
Db	1	MNPNSLEDFLRNVINKFHRALTLRETLQVIVEARIFLGVDRVITYKSPASDGSCEVLAE	60	
Qy	61	AVNRALSLSLGLHPRVEDIPQAREELGNQKMTAVDVANRRKKSHELSGISTESHN	120	
Db	61	AVNRALSLSLGLHPRVEDIPQAREELGNQKMTAVDVANRRKKSHELSGISTESHN	120	
Qy	121	GHTTVDSCHIOYLILAMGVLSLTVPMODDOLMGIMAVHSHKPRPTEOEETALLSK	180	
Db	121	GHTTVDSCHIOYLILAMGVLSLTVPMODDOLMGIMAVHSHKPRPTEOEETALLSK	180	
Qy	181	EVSIAITQSLSROYHQOQVDEALVQRETTVAQYGDREPTWQVLAETVGAVEADGAVL	240	
Db	181	EVSIAITQSLSROYHQOQVDEALVQRETTVAQYGDREPTWQVLAETVGAVEADGAVL	240	
Qy	241	YIAPDLTSSVAQHQWNIARFPMGNLFTSLNQLMRGQPSAAMEMAVQSTWEXPRPFT	300	
Db	241	YIAPDLTSSVAQHQWNIARFPMGNLFTSLNQLMRGQPSAAMEMAVQSTWEXPRPFT	300	
Qy	301	SVAPLPPTNCVPHGTYTLELBOBSDMIAPPELSAENQSFLLVPLADQQVGSLLILR	360	
Db	301	SVAPLPPTNCVPHGTYTLELBOBSDMIAPPELSAENQSFLLVPLADQQVGSLLILR	360	

```
QY 361 KEKSLVKHWAKRGIDRRNIIPLRSLFEAMEETOKLVPTWNSERKLAQVASTOLYMATIQ 420
D 361 KEKSLVKHWAKRGIDRRNIIPLRSLFEAMEETOKLVPTWNSERKLAQVASTOLYMATIQ 420
QY 421 QFVTRLTQQQAYAPLPTQVPMWIIIFNRQLTALLDALLEGKMWGVLYVAMDFKINSF 480
D 421 QFVTRLTQQQAYAPLPTQVPMWIIIFNRQLTALLDALLEGKMWGVLYVAMDFKINSF 480
QY 481 GHKRGDGLQEVADRLNOKLSPLAASPLLSRMHGDGFTILLTQISDQEMIPLCERILS 540
D 481 GHKRGDGLQEVADRLNOKLSPLAASPLLSRMHGDGFTILLTQISDQEMIPLCERILS 540
QY 541 TFOEPFLOQOPRYLTASMGISTAPYDGETAESLLKFAEIALTRAKCOGKNTYQFYPQD 600
D 541 TFOEPFLOQOPRYLTASMGISTAPYDGETAESLLKFAEIALTRAKCOGKNTYQFYPQD 600
QY 601 SAPMLDRLTESDLRQALTNQEFVLYXPOVALDQKLLGYEALVRMHPRLGOVAPV 660
D 601 SAPMLDRLTESDLRQALTNQEFVLYXPOVALDQKLLGYEALVRMHPRLGOVAPV 660
QY 661 IPLAEELGLINHLGOWVLETACATQHOFRETGRRLRMAVNISARQFODEKWLNSVLECL 720
D 661 IPLAEELGLINHLGOWVLETACATQHOFRETGRRLRMAVNISARQFODEKWLNSVLECL 720
QY 721 KRFGMPEDLELEITESIMEDIKGTVLLHRLREGEVQVALDQGTGYSLSILKQDPI 780
D 721 KRFGMPEDLELEITESIMEDIKGTVLLHRLREGEVQVALDQGTGYSLSILKQDPI 780
QY 781 HRLTIDSFVNDLNEGADTAIOYVIDLANGLNLETAEGIESAOQORLOKMGCHGQ 840
D 781 HRLTIDSFVNDLNEGADTAIOYVIDLANGLNLETAEGIESAOQORLOKMGCHGQ 840
QY 841 GYFLTRPLPABAMMYLYXPOILDFGPPPLPKYALPETETAGOGNDRPLPSLRE 900
D 841 GYFLTRPLPABAMMYLYXPOILDFGPPPLPKYALPETETAGOGNDRPLPSLRE 900
QY 901 NPWTEKLDYVLLERLOQRNVKEKLYLKANKIRASININDIIXSTVEYRQFLNDRV 960
D 901 NPWTEKLDYVLLERLOQRNVKEKLYLKANKIRASININDIIXSTVEYRQFLNDRV 960
QY 961 VLFENSGOWSVTESHNDSCSTINDEIDPCFKGHYLRXGRRAVSDIEKALA 1020
D 961 VLFENSGOWSVTESHNDSCSTINDEIDPCFKGHYLRXGRRAVSDIEKALA 1020
QY 1021 DCHKEILRHQYOKANLYPVVFENELMGLLTAHECKTPRYQOEEDLOLMELATQVAT 1080
D 1021 DCHKEILRHQYOKANLYPVVFENELMGLLTAHECKTPRYQOEEDLOLMELATQVAT 1080
QY 1081 HOGELYEOLLETANIRLOOISSLDALTQVGNRYLFDSTLEREMORLQIREPLALLCDVD 1140
D 1081 HOGELYEOLLETANIRLOOISSLDALTQVGNRYLFDSTLEREMORLQIREPLALLCDVD 1140
QY 1141 FFKFENNYGHPADRCRCKTADAMAKYAKRPDLYARYGGEERATILSETSLGAIWVT 1200
D 1141 FFKFENNYGHPADRCRCKTADAMAKYAKRPDLYARYGGEERATILSETSLGAIWVT 1200
QY 1201 EALOVEANALAPHTVSGTGHVTLSTIGIAYVTPERHINPNALVKAADLALTEAKAKGNQ 1260
D 1201 EALOVEANALAPHTVSGTGHVTLSTIGIAYVTPERHINPNALVKAADLALTEAKAKGNQ 1260
QY 1261 WLAYEGSOLPHVGEV 1276
D 1261 WLAYEGSOLPHVGEV 1276

RESULT 2
AD1953
hypothetical protein al11175 [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AD1953
```

```
R:Kaneke, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriju
Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD1953
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-880 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA873132.1; PID:g17130522; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: al11175

Query Match 14.6%; Score 967.5; DB 2; Length 880;
Best Local Similarity 36.4%; Pred. No. 8.56-54;
Matches 248; Conservative 108; Mismatches 232; Indels 93; Gaps 16;

QY 188 QSLSRQYHQOQVDEALVQRLTETVAOYGDRPETW-----QYALETGOAVEAD 236
D 261 OKQAEELHOR-----DRLLQAVAEAN-----WLVEVMYDRAIEKALAVGGAQAD 309
QY 237 GAVLY--IAPDLTGSVAQHYQMNLRFDW-GWLETSL--NQBIMRGOPSAMEPMAAYQS 291
D 310 RAYLFKNHHPHITGRRAV---SLOFEWYGSDDLPSIDQWONQD-----YQS 352
QY 292 TWEKRPPTSVAAPLPPTNCVPHGYTLCEIQRSDMIAPESLSAENFSPLTVPPLAADQ 351
D 353 T-ELSRKWSVLS-----SGKSINELRRKL-VAERELLENGIOSLLVPLGLEDO 401
QY 352 WVGSLILLRKEKSLVKHWAKRGIDRRNIIPLRSLFEAMEETOKLVPTWNSERKLAQV 411
D 402 FMGYLGIA-----DQYERY--WSRHEESTLITMA 429
QY 412 TQLYMATIQQVTRLTQQQAYAPLPTQVPMWIIIFNRQLTALLDALLEGKMWGVLYVAM 471
D 430 ASISGAMORQOVEEKIRYQALHDLTLGLPNRLLENELAKALANASQESLAVWFLLD 489
QY 472 RFRKRSFHKGTGGLQEVADRLNOKLSPLAASPLLSRMHGDGFTILLTQISDQEM 531
D 490 RPKVYNIDLGHITLGBRLLQSVQNRRLDCLRS-----GDVYSWVGDEFTILLPHSYIEV 545
QY 532 IPLCERILSTFOEPFLOQOPRYLTASMGISTAPYDGETAESLLKFAEIALTRAKCOGN 591
D 546 TQACRRIKALAEVEFHLDGHELYASASLGIALDLHNSPDALILKHADAALYHAKDEGRN 605
QY 592 TYQFYRPQDSAPMLDRLTESDLRQALTNQEFVLYXPOVALDQKLLGYEALVRMHPR 651
D 606 NYQFTYTSLSGKTPBELTLEKSLRVALEKQELKAYQYQVRNVLITGQITGMALRMQHP 665
QY 652 LGQVAPDVFILAEELGLINHLGOWVLETACATQHOFRETG-RRLRMAVNISARQFODE 710
D 666 MGIYAPSFITFIABETGILIPIGEMVLTACOMQR-AMQEGALRPYQAVANLSKQFQRP 724
QY 711 KWLNSVLECLKRTGMPEDLELEITESIMEDIKGTVLLHRLREGEVQVALDQGTGYS 770
D 725 TLLEIVSAIILAETGLAEFLELEITESPAIEDIFTRSVRLNLOOMGVADIIIDFGTGS 784
QY 771 SLTLKQPIHRLTIDSFVNDLNEGADTAIOYVIDLANGLNLETAEGIESAOQOR 830
D 785 SLSRQLPLPLNLKIDSGFIALTTDSKVAHIVAKAIVTLGRSLDLRTLBAEVEKQEBDF 844
QY 831 LQKMGCHLGOGYFLTRPLPAE 851
D 845 LKSTINCQDVQSGFLFKPLSAE 865

RESULT 3
C83070
conserved hypothetical protein PA4601 [imported] - Pseudomonas aeruginosa (strain PAO
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
```





C:Accession: S76102  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O. K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*  
S.  
A:Reference number: S74322; MUID:97061201  
A:Accession: S76102  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1244 <KAN>  
A:Cross-references: EMBL:D63999; GB:AB001339; NID:g1001396; PIDN:BA10080.1; PID:d101073  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 13.1%; Score 865; DB 2; Length 1244;  
Best Local Similarity 31.0%; Pred. No. 5,9e-47;  
Matches 259; Conservative 141; Mismatches 286; Indels 150; Gaps 28;

QY 145 VPVMDQ-----QLMGIMAV-----HHSKPRRF-----TEDEWETMAL----- 177  
DB 475 IPFVADQATEITLPLDGSNVAEMKRVSPIMQKGAFLVSLRDITEQHOARLALAESPK 534  
QY 178 -----LSKEYSLAITSQSROYHQ-----QVQELVQRLTTYAQYGDREPTWQ 223  
DB 535 KYRIVELTSGEIMWILDDQDTTFANQOLADMLGYVQELKEKNTAFVLYIHMLPEESON 594  
QY 224 YALETV-----GOAVEA-----DGAVALYAPDLTSGVAOHYQGNLRFQGNW-- 265  
DB 595 SHOKTLOSPRCVLPNHNQYDVQQRDRGSYLV-----GLVRSANTWD---QMGNTRG 645  
QY 266 ---LETSIMQELMRCQPSAMEP-----MAAYOS--TWKRPRTSVAPLPTNCVPHGYT 316  
DB 646 ELAMLTDTIKKRSASQALASBQRLGILGSIQDVVMSADAVSFATLYLPTTAMVYQOS 705  
QY 317 LGELEQRD-W---IAPESLSAENFQSLVPLPLADQOWGSLILKREKSLVYHMAK 372  
DB 706 LEVCYQSNFWEQYHNPDRLLLEYHQLL---MERDO-----TELEKRIYVPGSKE 754  
QY 373 RGIDRRNLPR-----LSFEAMETQKIVPTWNSERKLAQVASTQLYMATIQOFVTRL 426  
DB 755 RWLPRRSQLVADGHEQRLRIDSIDS-----DITERKLA-----AEK 790  
QY 427 ITQOTAYDPLTQLPWITFNQPLTALLDALYEGKM--GVULVIANDREKRINESFGHKTG 485  
DB 791 LHYANAHSDITNLNPRSMFLDRIGHALORNLRRDLRPAVLFLDGDGKIIINDSGHSCG 850  
QY 486 DGLQEVADRINOKSLPLAAYSPLLSRWNGSGFTILLQISNOEMIGLCEHLLSTPQER 545  
DB 851 DLLQGGIADRRLRQCLRP-----EDTLARLGSGDEFTMLNITCPEDEVIANVQRIHOELQRP 906  
QY 546 FFLQGPRIYLTASMGIS--TARYDGETASLLKFAEIALTRAKCOGKNTYOFYRPODSAPM 604  
DB 907 FNLNGQEIFTMTSIGIALNHPHYGH--PDQVLRADDTAMYRAKAKGKGYALFNQGMHNA 965  
QY 605 LDRLTLEDLQALTNQFVLYFQPOVALDITGKLGVALYRMQRPRLGOVARDVFIPLA 664  
DB 966 VQRLQRENDLRRALDRLOLQHYQIVCLKTQLOGVALYRMQRPREGGLILPREFGAIA 1025  
QY 665 EELGILNLGQWVLETA-----CATHQHFRETGRRLRAVNVISAROFODEKMLVNLCL 720  
DB 1026 EETGLIYVMQWIMLEASROLLELKQSPOLS--HLQVSIWSSQQLDORLKLKIVDEL 1083  
QY 721 KRTGMPEDLELETESLIMEDIKGTVLLRLHREGEQVVALIDFGTGYSSLSILKQDPI 780  
DB 1084 SSTNLAPDQDLKETESLILIONLNLAAOVLSLRQNTQISLDGCTGYSSLSYLRHPI 1143  
QY 781 HRLKIDKSFVVDLNEGADTAIIQYVIDLANGLNETVAGEIGESAOLORLQKMGCHUGQ 840  
DB 1144 NTIIVDRSFVVTMEPNNTAIIVHTIYTLAHTLIGDVAIEGIEFERHNLTOHMLGCCDAQO 1203  
QY 841 GYFLLRPLPAPAMMYLYLYPQILDGCPRLPKVALPETEAGGNGNDRLPLRS 896  
DB 896 GYFLLRPLPAPAMMYLYLYPQILDGCPRLPKVALPETEAGGNGNDRLPLRS 896

DB 1204 GYFARPIPSEDLDFL-----TIQIPMSVPAN-TECH-----GDRKEPPS 1244

# RESULT 6

Conserved hypothetical protein PA0575 [imported] - *Pseudomonas aeruginosa* (strain PAO H83574)

C:Species: *Pseudomonas aeruginosa*

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: H83574

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; L.

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pa

A:Reference number: A82950; MUID:20437337

A:Accession: H83574

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1245 <STO>

A:Cross-references: GB:AE004493; GB:AE004091; NID:9994636; PIDN:AG03964.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA0575

Query Match 13.0%; Score 862.5; DB 2; Length 1245;

Best Local Similarity 29.1%; Pred. No. 8,5e-47;

Matches 263; Conservative 154; Mismatches 345; Indels 141; Gaps 24;

QY 20 RALTRERLOVYVEARFLGVDRKITYKFS-----DSGEVLA----- 59

DB 412 RDLRLR-GIMIDITEAKLTEQAMRLESEKGFASVFNHCPDMVYLANGADGRELAVNSTEQ 470

QY 60 -----EAVNRALPLSLGHFVEDIPQARELGNQRMKIAVDVAHRRKSHLSGR 112

DB 411 QIGIADDAIGKTA--TELGI-WALPGIGPQVLEKLGH-GNLTNIEVLRNRNGSTFSLA 526

QY 113 ISPTENSHGHYTV-----DSCHI--QYLLAM-----GVLSLTVPVMDQOL 153

DB 527 LS-AQHVALLDQPALVAVIRDTIHLVETOELLRISEKFANAFHAYPGGLISLSDGTLL 585

QY 154 WGIAMVHHSKRRRFRTEQEMETALLSKFVSLAITSQOLSROYHQOQVQELVQRLTTYA 213

DB 586 ---IDVNGEFTRLNGYRDEVYISRSTLELGLWVDETDRKRLISLVR-HHTLQGFAYVR 641

QY 214 QYGDREPTWOYALETVGOAVEADGAVLYIAPDLTSGVAOHYQGNLRFQGNWLETSMQE 273

DB 642 DRNGITRQCEMSAHHI--SIDGEDCVLTARDIT-----ERQLMQE 680

QY 274 LMRGQPSAMEPMA---AVQSTWEK---PRFTSVAPLPTNCVPHGYTLGELQNSDW 326

DB 681 KIQ-QAATVFESTAGVMTIDTQRITAVNRAFSKIT-----GYS---EOEALG 725

QY 327 IAPPSLSAENQSFLLY---PLAADQOWGSLILKREKSLVYKHWAKKRGIDRRNLPR 383

DB 726 RSPSLSSGQSDSFFLYLMMQNLDERDGHQGEINRRKKTGTGELPYM----- 771

QY 384 LSFEAMETQKLVPTW-----NRSEKRLAQVASTQLYMATIQOFVTRLITQOTAVDPLTQ 438

DB 772 LTISAVNHQGEITTFHGVGFADISTLKVAQAR-----LDYQAHNDPLTG 815

QY 439 LPMNITFRKQLTALLDALYEGKMGVLYIAMDREKRINESFGHKTGGLGLOEVAADRINQ 498

DB 816 LPNRLIFESRILNHALDEARESSRGAVLFIIDLDRKHINDSLIGHITGLLAKAIAERLDR 875

QY 499 KLSPLAAASPLLSRWNGSGFTILLQISDNQEMIPLCRSLSTPOEPFFLOGQPIYLTGA 558

DB 876 QLRDV---DTVARLGGEFTILLPGHQBESDAEHVARKLNLATFAPQADGHEFFVSAS 931

QY 559 MGISTAPYDGETASLLKFAEIALTRAKCOGKNTYOFYRPODSAPMLDRLTLESDLQAL 618

DB 932 VGIALFPKDGDAPIYLVNNAQAAMYRAKSRGRSRIEYTRRELYTLATBRMALLETIRLAL 991

QY	619	TNDEVLXFEQVVALDMDGKLLGEALVYRMQHPRLGQADVFPLAEEGLNHLNGQWL	678
Db	992	ERDELQTLQYQKLSLESGLLVGAEALRWYHNPJLGEISPERFPLAEDGGLLPJGDWL	1051
QY	679	ETACATHOFFERENGRLRMAVNISAROFODEKMWLSVLECLKRTGMCPDEDELEITESTL	738
Db	1052	EHACQOMQEMQKTHAPGRLPSVNLAGQIQPQLIERLQELQESLSEPSRQLQLEITESP	1111
QY	739	MMEDIKGVVLLHRLREEGVQVALIDPGTGYSSLSILKQLPIHRLKIDKSFVNDLMEGA	798
Db	1112	IMNQTLEALVHLKRLRGVQLAIDPGTGYSSLSYLKRLPLDILKIDKSFIRGLPDOPH	1171
QY	799	DTAIIQVYIDLANGLNFTVAEGIESEFAQLORLOKKMGCHLGQGFELTBPAPAMNTYLY	858
Db	1172	DAATRAIRAIAGRSMQGLTVIAEGVEYEGQGSFLTHGCGIQGFSVLPSPPLPAELASKFL	1233
QY	859	YPO 861	
Db	1232	KPR 1234	

RESULT 7  
AB2334  
hypothetical protein all14225 [imported] - Anabaena sp. (strain PCC 7120)  
C:Species: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120  
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_revision 11-Jan-2002  
C:Accession: AB2334  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuriat, T.; Sasamoto, S.; Watanabe, A.; Iriuchika, N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AB2334  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1565 <KUR>  
A:Cross-references: GB:BA000015; PIDN:BA875924.1; PID:917133360; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: all14225

```

Query Match      12.9%: Score 836.5; DB 2; Length 565;
Best Local Similarity 40.4%; Pred. No. 5.4e-47;
Matches 193; Conservative 91; Mismatches 171; Indels 23; Gaps 4.

QY 381 LPLSPEAMEETQKIVPTWNSERSLQAVASQVLMATIQOEVTRLITQOTAVDPLOTP 440
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 102 LPSLPEFVEDEGTLMDTSTQTLNKLDE-----LIDYISNWKTLGTP 144
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 441 NMIIENROTLTLLDALYEGKNVGLVIMAFRRKRNESFGHKTGIDGLOEVAIDLNOKI 500
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 145 NBEILBERNQA-LSQOYSQKMLAVMYGLIDDPFTGISAHLEBQSULLIRAVAQRLTSL 203
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 501 SPLAAYSPLSRKMGHGGFTLLITLTQISDNQEMTPLCEIRLLSTFQEPFFLOGEPITYLTASMG 560
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 204 ----AKQDILHLSDEDFEARVEIPSIETIKLSQLLTLTNKPTTGQNIHTASIG 259
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 561 IS-TAPYDGETAESLLKRAEIALTRAKCOGKNTYQFRRQDSAPMIDRLTQESDRLQAT 619
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 260 ITINQPDNIHDVDDLLQQAHVALLYAKQOGRSQHOFYSPETNAQOERLATELENLGALE 319
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 620 NOEFPLVFOVALYTGKLLGVEALVVRMCHPLTGOVAVPFLPLAEELGLNHLHGOVLE 679
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 320 RNEIIVYVYPIIDLOTKOTTAVEALVVRMCHPRTGLVSPAKPTPLAEANGLIVEIGEWVR 379
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 680 TACATQHQHFRETGRRLMAVNIISARQFODEKMLNSVLECLRTGMPDELELELEISIA 739
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 380 TACIQNRNTQMLGFLPIRISVNLISARQFESNMIVELYSILIDESGHPRIYLELVETESSL 439
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 740 MEDIKGTVLLHRLREGVOVAIIDQGTGYSSSLITIKOLPIRLKIDKSFVNDLNEGAD 799
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

Dbb 440 MTDIDRSVTIIKQLRELQELGVMIALDDEGTGGTGYSLNLYKREFYNNMLKIDRSFYQDVCSNPDS 499

Oy 800 TAIIDYVIDDLANGLNLETVARGISIEAOLORLQMGCHLGCGIFLTRPLPAEAMMTYL 857  
| : | | | | : | | | | | : | | | | : | | : |  
Db 500 AAVTAIITLALKSLQDKTTAGSIEHQOOLSTLRKRCGEQGGYFFGLPAPAKETMELL 557

RESULT 8  
EB3610  
conserved hypothetical protein PA0285 [imported] - Pseudomonas aeruginosa (strain PAO  
C.Species: Pseudomonas aeruginosa  
C.Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C.Accession: EB3610  
R.Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Iarbig, K.; L  
Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A.Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
A.Reference number: AB2950; MUID:20437337  
A.Accession: EB3610  
A.Status: preliminary  
A.Molecule type: DNA  
A.Residues: 1760 <SIO>  
A.Cross-references: GB:AE004466; GE:AE004091; NID:99946120; PIDN:AA03674.1; GSPDB:GN  
A.Experimental source: Strain PA01  
C.Genetics:  
A.Gene: PA0285

Query Match	12.6%	Score 830.5	DB 2:	Length 760:	
Best Local Similarity	33.3%	Fred. No. 4.2e-45			
Matches 238:	Conservative 104:	Mismatches 248:	Indels 125:	Gaps 18:	
QY	214	QYGDPRPFWQVLA--ETVQAQAEADGAVLYIAPDITGSAVQAHYQWMLRPDMGN--WLETS	269		
DB	120	ELGDPLETFEKKRLHREEVATVLEARNHL---QGLTDNDLHYR--LRKKDDYKMHIS-	173		
QY	270	LMQELMRGQPSAMERMAAVOSTWEKRPREFTSVAPLRPLNNVCYPHGYTIGLEQO-----	322		
DB	174	---RGR-----VLRDALKRPLHYTGVA-----RDITLQRLKEDHLRQA	209		
QY	323	-----RSDMI---APPESLSAENPQSF---LIV	344		
DB	210	AVFDSTRBEVLTDAQAVLYHVNPSFERITGTGRSEVYLKTAILLRSGQDQAFQRYML	269		
QY	345	PLAADQWVGSLLLRKEKSLVKHNAKGRGIDRRNITPLRSEAMEETOKLYPTWNRSE	404		
DB	270	ALREDQVMSGEIWMNRKSGEIVPQWHLIRAY--RNDQGLTHYGVGFSQ--LSSIKRSD	325		
QY	405	KLAQVASTQLYMAITQQFATRLITQOTADPLTQLRPNMILFNKQLTALLDLLEGKMG	464		
DB	326	ELDFLAH-----HDSLTGPRNVLRLRERIEQALENKDKRTLVAGA	364		
QY	465	VLVIAMDPRKIRISFGHKTGDLLQEOVADRINOKSLPLAASPLIRSNHNGGFTLITLQ	524		
DB	365	LLILDLDHKKHINDSLGHTTGDMILKEVSKRKHQHDENC---LLSRIGSGEFALV-E	419		
QY	525	ISDNOEMTDLCERLISTFOEPFLLOGQRIYLTASMGISTAPYDGETAESLKFPAETALTR	584		
DB	420	NDDPEFAVALSORIILDFGNAPFDIHCQPIYISASIGSVSLPEDASDVYDHLMOHADAALFQ	479		
QY	585	AKCCKKNTPQAFRRQDSAMIDLITLESRLQALNNOEVLLEFORQVALDRCKLLGVEVL	644		
DB	480	AADSGNNATFAFTRVLTARAHVQESALRHALDELDRLAYHOPRHDLASGRIVGESL	539		
QY	645	VMQHPRLQGVAPDFETPLAEELGLNLHNGOAVLETACATNHFHFETGRTRRYR--MAVNS	703		
DB	540	VMQHPRERGLVPRGEFVPAVEECGLIADLNNVLRKACRQMBE--MQQRVELEFPAVNS	598		
QY	704	AAQFQDEKKLNVLECKIKTGMPPREDLELETESLMEIDIKCTVULHRLREGQVAID	763		
DB	599	SHLFNNGLEERTIANALESSGLEPRYLELEVESVAMDEQDSLNLCLCRITLGVNAD	658		



DB 745 EGVETAEQLTYLQSLHCHDEVQGFIAKPMPEVAL 778

## RESULT 11

AC1834  
hypothetical protein al10219 [imported] - Anabaena sp. (strain PCC 7120)

C:Species: Anabaena sp.  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002

C:Accession: AC1834

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kunitz, T.; Sasamoto, S.; Watanabe, A.; Iritiguchi, Nakazaki, N.; Shikama, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A:Title: Complete genomic sequence of the filamentous nitrogen-fixing Cyanobacterium Anabaena

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AC1834

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-585 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA07743.1; PID:q17135197; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: al10219

C:Superfamily: hypothetical protein s111895

Query Match 12.3%; Score 813.5; DB 2; Length 585;  
Best Local Similarity 35.1%; Pred. No. 3.3e-44;  
Matches 199; Conservative 107; Mismatches 184; Indels 77; Gaps 11;

OY 314 GTTIGLEBRS--DMIAPESSIAENFOSFLVPLAA-----DOOW-----VGS 335  
DB 57 GKPEELVKNMFDEFHPDITSNSQSFNYLIONGATVLVESFRFRKSGSMCFLESTGS 116  
OY 356 LILIRKE-KSLVKNHMAKKGIDRRNITLPLSFEAMEETQKLVPTNRSEKLAQAVASTOL 414  
DB 117 NLAEPSYKGV-----INSRD-----TRKLA----- 141  
OY 415 YMAITQOEYTRILITQOTAYDPLTQLPNWIIFNRQLTALLDA-LYEGKMWGLVVIAMDR 473  
DB 142 -----ERLV-HDALHDVLTGLPNRVLTIDRLRAVEYAKRYSDDLFAVLFLDLDR 191  
OY 474 KRINESFGKRTGDLQEVADRNLCKSLPLAVSPILSRMHGDFITLLTQISDNQEMIP 533  
DB 192 KRINSLGHTIGDOLLVITIAORLIECLRP-----TDIAARFGDEFTILLEGIODISDVR 247  
OY 534 LCERLSTFOEPFLOGPIYLTASWGISTAPYDGETAELIKFAETALTRACOGKNY 593  
DB 248 VVERIOEKLIVVLSGHEITFTASIGLSATGYQPEDLNRNADIAMRAKARACT 307  
OY 594 QFYRPDASAPMLDRLTLESDLQALTNQEFVLVFOPOVALDTGKLLGVEALVWQHPRIG 653  
DB 308 EIFNSDMHVOIVERLOLENDRLRAIRHRELVYQPIVSLTGRTGFEALVWMLPEQG 367  
OY 654 QVAPVFTIARIELGILNLGOMVLEACATQHOFREI-----GRRLAAVVISAR 705  
DB 368 IYFPEFMIAQETGLIPIDAMVLEACRQTRQWQEQIPSLSTDLAKPLSISINLCS 427  
OY 706 QFODEKMLSVLECKLRTKMPPEDELEETESIMMEDIKGTAVLLHRLREEGVOVADIP 765  
DB 428 RSGSKLLEDINOVLODITDLDAOSLKELETESYIMENGENATMLQRLNLGIELIDDE 487  
OY 766 GTSQSSSLTKQLPIHRLKIDKSFVNDLLNEGADTAIIQIVYIDLANGLMEVAGIESE 825  
DB 488 GTGSSSLGRLHMFPIINGLKIDQSFVSGRVEAGNLIHVEIYVTLSSKLGDVTAEGVEIP 547  
OY 826 AQLQRLQKMGCHLGGCYFTRPLPAEA 852  
DB 548 AQLQRLRELKCEYGGGFSEPLSDS 574

RESULT 12

B83232  
conserved hypothetical protein PA3311 [imported] - Pseudomonas aeruginosa (strain PAO)

C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: B83232

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L.

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: A82950; MUID:20437337

A:Accession: B83232

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-783 <STO>

A:Cross-references: GB:AE004753; GB:AE004091; NID:g9949433; PIDN:AAG06699.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA3311

Query Match 12.3%; Score 811.5; DB 2; Length 783;  
Best Local Similarity 29.7%; Pred. No. 7.3e-44;  
Matches 239; Conservative 119; Mismatches 285; Indels 161; Gaps 19;

OY 160 HSKPRPF-----TEQEWETALLSKE-----VSLATQSQLSROYHQ 198  
DB 43 HPRRPVRFVQGDPTAMOGRLFESPDVGYVLONCTYSPSLAFLVACLAGYTALD 102  
OY 199 QVGEALVQRETTVAQYGDPRFTWQ-----YALETV 229  
DB 103 -----MERGNSLS-----HPLMOWIGAFCLSGIWAHIFVAMFAHADIARIDYLPIT 153  
OY 230 GQAVADGAVALYIAPDLT-----GSVAQYQWMLRFDMGNWLETSL 270  
DB 154 GSLILAVASVLYFMWTATAPRFGLLPCLLAACCIGIATAMHY----- 197  
OY 271 WOELMRGQPSAMEPMAVAVOSTWEKP-----RFTSVADLPPTNCVPHGYTLG 318  
DB 198 -----TGMAAMRSVATQYQPSLFLSVLIAIGAFTALAAVP-----YLRG 239  
OY 319 ELQRSDMTAPRPSL-----SAENFQSF-----LIYPLADQQWGSLLIRREKSLYKH 368  
DB 240 RRSARTRYMKLINSLLAGIAIMHFTGMALVAVSPAGT-----PLEQASNDLSRLG 293  
OY 369 WACKRGIDRRNITLPLSFEAMEETQKLVPTNRSEKLAQAVASTOLYMAITQ-QFYTRLI 427  
DB 294 WL--TGVLASAIACGIWAMSEKQ-----ERLSE--NSRVALLNQDHNHASTL 341  
OY 428 TQOTAYDPLTQLPNWIIFNRQLTALLDALYEGKMWGLVVIAMDRKRINESFGKRTGDG 487  
DB 342 ROMARYDSLTGLONRTAFNEVFVOHLENCRLRGKGLAVMLDLDDHFRKINDSLGHSQDQ 401  
OY 488 LLOEVADRNLQKSLPLAAYSPLSRMHGDFITLLQISDNQEMIPCELCSTFOPEPF 547  
DB 402 LKTIYSEIRISVLRD-----SDVAFRAGDEFCVALIDTQDHEHT--LSQRLMKMKEPYA 436  
OY 548 LOGOPITLFAWSMISTAPYDGETAELIKFAETALTRACOGKNYQFYRPDASAPMLDR 607  
DB 457 LDRRTLVMTASVSVSLYPNNGQCEELKKNAGLALHQSKCGKNNQGFRRQLVLRATQE 516  
OY 608 LTLESDLQALTNQEFVLVFOPOVALDTGKLLGVEALVWQHPRIGVAVDFIPLAEEL 667  
DB 517 LQMEELRQALRDQDOLELHVOPTIALADGEVHOLEALVBRHPTQGLIGPDRFIGLAEAN 576  
OY 668 GLINLHGOAWLETACATQHOFREIETGRRLMAVNISARQDQDEKMLNSVLECKLRTGMP 727  
DB 577 GMTDQDDWVLRACRDLRLHLAGHERLVAVANCCASNIGRASIVDEVRHNALEQGLAA 636  
OY 728 EDLEETTESIMMEDIKGTAVLLHRLREEGVOVADDFGTGSSSLTKQLPIHRLKIDK 787  
DB 637 CFLELEVTEDALMYNNDQITPLERLRELGVSLSIDDFGTGSSSLAVLRRLPLDALKVDR 636



[illegible]

```

Db 402 TRSNRRYFYIASVDSSEMRRRRELEKDLROLQJNHELHLYUQPVUDYRDHRRVGVYALLR 461
QY 647 WOHPRLGQVADVFETPLAEBLGLINHGQVWLETACATHOHFFERTGRRLMAVVISARQ 706
Db 462 WOHPRHGFVPDPDETFPLAENGSIFFSIGEWLDDOACROREMHDGDDDFRMAVNISTVQ 521
QY 707 FODEKWLNSVLECKRKGMPDEDELETESLIMEDIKGTIVLLHLRREGGVQVATDDFG 766
Db 522 LHHNAPRVVSNLQVYKRLPARSLELEVTETGLMEDISTAAQHLISLRAGALAIADDFG 561
QY 767 TGYSSLSILKOLPILHRLKIDKSPFNLDNEGADPAIIQVYIDLANGILETVAGIESEA 826
Db 582 TGYSSLSTLKSPLPDKIKIDKSFQODLLODDBDKATYIRAILIQLOKSLGMQVINGVYTAE 641
QY 827 QLORLQKMGCHLGGCYFLTRPLRPAMMTYU 857
Db 642 QEAYIIAEGCNEGGOGYLSKPLPARLTOYL 672

```

Search completed: June 7, 2002, 18:56:32  
Job time: 206 sec

RESULT 15  
F83100  
Conserved hypothetical protein PA4367 [imported] - *Pseudomonas aeruginosa* (strain PA01)  
C:Species: *Pseudomonas aeruginosa*  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
R:Stoever, C.K.; Plam, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bardenheer, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
A.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen  
A:Reference number: AB2950; MUID:20437337  
A:Accession: F83100  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-687 <STO>  
A:Cross-references: GB:AE004852; GB:AE004091; NID:99950587; PIDN:AGC07755.1; GSPDB:GN00101  
C:Genetics:  
C:Gene: PA4367

[illegible]

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: June 7, 2002, 18:54:01 ; Search time 19.14 Seconds

(without alignments)  
2581.303 Million cell updates/sec

Title: US-09-272-809-2

Perfect score: 6614

Sequence: 1 MNPRLSLEDPLRNYINKFHR.....GRNOMLAYESQLPHVGEV 1276

## Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6614	100.0	1276	1	PHY2_SYNY3
2	821.5	12.4	696	1	Y195_SYNY3
3	708	10.7	827	1	Y41L_RHISN
4	684.5	10.3	799	1	YDDU_ECOLI
5	621	9.4	623	1	YD54_MYCTU
6	607	9.2	661	1	YCTR_ECOLI
7	561	8.3	651	1	YHUK_ECOLI
8	535	8.1	735	1	YNTC_AZOCA
9	400	6.0	528	1	YVCC_ECOLI
10	386	5.8	782	1	YVIE_ECOLI
11	370	5.6	747	1	YRGE_ECOLI
12	368.5	5.6	307	1	YD57_MYCTU
13	366	5.5	516	1	YLAB_ECOLI
14	361	5.5	507	1	YCGG_ECOLI
15	358.5	5.4	532	1	YOAD_ECOLI
16	356	5.4	518	1	RTN_ECOLI
17	354.5	5.4	1000	1	Y041_SYNY3
18	323	4.9	362	1	YAH4_ECOLI
19	296.5	4.5	753	1	YFEA_ECOLI
20	290	4.4	729	1	YFEA_ECOLI
21	283.5	4.3	765	1	YFAY_ANASP
22	280	4.2	452	1	YCDT_ECOLI
23	273.5	4.1	751	1	YRGE_ANASP
24	265.5	4.0	1105	1	YRGE_ECOLI
25	265	4.0	371	1	YALC_ECOLI
26	262	4.0	1129	1	YRGE_ECOLI
27	261	3.9	1115	1	PHYB_PHLNU
28	258	3.9	1124	1	PHYB_PHLNU
29	257.5	3.9	748	1	PHYB_PHLNU
30	249.5	3.8	1136	1	PHYB_PHLNU
31	247	3.7	1132	1	PHYB_PHLNU
32	244	3.7	1124	1	PHYB_PHLNU
33	242.5	3.7	728	1	PHYB_PHLNU

34	237.5	3.6	1131	1	PHY_PINSY	Q41046 pinus sylve
35	237.5	3.6	1132	1	PHY_PINSY	P36505 physcomitre
36	236.5	3.6	1132	1	PHY_PINSY	P54595 bacillus su
37	236	3.6	1124	1	PHY_PINSY	P15001 pisum sativ
38	235.5	3.6	1137	1	PHY_PINSY	Q92419 oryza sativ
39	234.5	3.5	1112	1	PHY_PINSY	P42498 arabidopsi
40	233.5	3.5	1172	1	PHY_PINSY	P14713 arabidopsi
41	231.5	3.5	1121	1	PHY_PINSY	Q39557 ceratodon p
42	231	3.5	646	1	PHY_PINSY	P13518 escherichia
43	230.5	3.5	1128	1	PHY_PINSY	P10931 oryza sativ
44	229	3.5	1124	1	PHY_PINSY	P06592 cucurbita p
45	226	3.4	1118	1	PHY_PINSY	P42496 adiantum ca

## ALIGNMENTS

Result ID	PHY2_SYNY3	STANDARD:	PRT: 1276 AA.
AC	Q55434;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	01-MAR-2002 (Rel. 41, Last annotation update)		
DE	Phytochrome-like protein cph2 (Bacteriophytochrome cph2).		
GN	Cph2 OR SL0821.		
OS	Synechocystis sp. (Strain PCC 6803).		
OC	Bacteria: Cyanobacteria; Chroococcales; Synechocystis.		
OX	NCBI_Taxid=1148;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=96127529; PubMed=8590279;		
RA	Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,		
RA	Sugiyama M., Tabata S.,		
RT	"Sequence analysis of the genome of the unicellular cyanobacterium		
RT	Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb		
RT	region from map positions 648 to 928 of the genome.";		
RL	DNA Res. 2:153-166(1995).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=97061201; PubMed=8905231;		
RA	Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,		
RA	Miyajima N., Hirosewa M., Sugiyama M., Sasamoto S., Kimura T.,		
RA	Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nario K., Okumura S.,		
RA	Shimizu S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,		
RA	Tabata S.,		
RT	"Sequence analysis of the genome of the unicellular cyanobacterium		
RT	Synechocystis sp. strain PCC6803. II. Sequence determination of the		
RT	entire genome and assignment of potential protein-coding regions.";		
RL	DNA Res. 3:109-136(1996).		
RN	[3]		
RP	CHROMOPHORE 1 BINDING, AND MUTAGENESIS OF CYS-129 AND HIS-130.		
RX	MEDLINE=20435261; PubMed=10978170;		
RA	Park C.-M., Kim J.-I., Yang S.-S., Kang J.-G., Kang J.-H., Shim J.-Y.,		
RA	Chung Y.-H., Park Y.-M., Song P.-S.,		
RT	"A second photochromic bacteriophytochrome from Synechocystis sp. PCC		
RT	6803: Spectral analysis and down-regulation by light.";		
RL	Biochemistry 39:10840-10847(2000).		
RN	[4]		
RP	IDENTIFICATION OF TWO BILIN LYASE DOMAINS.		
RX	MEDLINE=20510044; PubMed=11063585;		
RA	Wu S.-H., Lagarias J.C.,		
RT	"Defining the bilin lyase domain: lessons from the extended		
RT	phytochrome superfamily.";		
RL	Biochemistry 39:13487-13495(2000).		
CC	-1- FUNCTION: PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT ARE		
CC	REVERSIBLY INTERCONVERTIBLE BY LIGHT. THE R FORM THAT ABSORBS		
CC	MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE FR FORM THAT		
CC	ABSORBS MAXIMALLY IN THE FAR-RED REGION.		
CC	-1- DOMAIN: Both bilin lyase domains bind with the bilin tetrapyrrole		
CC	chromophore precursor. The domain 1 shows red, far-red light		
CC	photo-reversibility. The domain 2 is photochemically inactive.		
CC	-1- PFM: CONTAINS TWO COVALENTLY LINKED TETRAPYRROLE CHROMOPHORES.		

```
CC -1- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D64003; BAA10536.1; -
DR Interpro: IPR001633; DUF2.
DR Interpro: IPR00160; DUF9.
DR Interpro: IPR003018; GAF.
DR Interpro: IPR001294; Phytochrome.
DR Pfam: PF00563; EAL; 1.
DR Pfam: PF01590; GAF; 3.
DR Pfam: PF00990; GGEF; 2.
DR SMART; SM00267; DUF1; 2.
DR SMART; SM00052; DUF2; 1.
DR SMART; SM00065; GAF; 3.
DR PROSITE; PS00245; PHYTOCHROME_1; FALSE_NEG.
DR PROSITE; PS50046; PHYTOCHROME_2; 2.
DR Photoreceptor; Phytochrome; Chromophore; Repeat; Complete proteome.
KW DCMAIN 1 197 BILIN LYASE DOMAIN 1.
FT DCMAIN 939 1075 BILIN LYASE DOMAIN 2.
FT BINDING 129 129 CHROMOPHORE 1 (PROBABLE).
FT BINDING 1022 1022 CHROMOPHORE 2 (POTENTIAL).
FT MTAGEN 129 129 C-S: HOLOPROTEIN EXHIBITS NO
FT MTAGEN 130 130 PHOTOCROMIC ACTIVITY.
FT MTAGEN 130 130 H-S: CHROMOPHORE LIGATING ACTIVITY (IN
FT MTAGEN 130 130 VITRO) IS 30-40% LOWER THAN WILD-TYPE.
FT MTAGEN 130 130 H-S: CHROMOPHORE LIGATING ACTIVITY (IN
FT MTAGEN 130 130 VITRO) IS ABOUT 10% MORE EFFICIENT THAN
FT MTAGEN 130 130 WILD-TYPE.
SQ SEQUENCE 1276 AA; 144686 MW; 318CF3A73962D99E CRC64;
Query Match 100.0%; Score 6614; DB 1; Length 1276;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 481 GHKTGDGLQEVADRLNKLSPLAAYSPLSRMHGDFTLLTLOISDNQEMIPICERLLS 540
DB 481 GHKTGDGLQEVADRLNKLSPLAAYSPLSRMHGDFTLLTLOISDNQEMIPICERLLS 540
QY 541 TPQEPFQGGPIYILPASMGSTAPYDETSLSLKFAEIALTRKCCGKNTYQFYRPQD 600
DB 541 TPQEPFQGGPIYILPASMGSTAPYDETSLSLKFAEIALTRKCCGKNTYQFYRPQD 600
QY 541 TFOEFFFLOGQPIYILPASMGSTAPYDETSLSLKFAEIALTRKCCGKNTYQFYRPQD 600
DB 541 TFOEFFFLOGQPIYILPASMGSTAPYDETSLSLKFAEIALTRKCCGKNTYQFYRPQD 600
QY 601 SAPMLDRLTLESDDLROALTNOEFVLYPOVALDPTGKLGLGEALVLRMHQPRLGQVAPVDF 660
DB 601 SAPMLDRLTLESDDLROALTNOEFVLYPOVALDPTGKLGLGEALVLRMHQPRLGQVAPVDF 660
QY 661 IPLAEELGLINHLGQWVLEACATQHFFRETGRRLRNAVNI SARQFODEKWLNSYLECL 720
DB 661 IPLAEELGLINHLGQWVLEACATQHFFRETGRRLRNAVNI SARQFODEKWLNSYLECL 720
QY 721 KRTGMPPEDELLETESLAMEDIKGTVVLRLRREGQVVAIDDPGTGYSLSLTKOLPI 780
DB 721 KRTGMPPEDELLETESLAMEDIKGTVVLRLRREGQVVAIDDPGTGYSLSLTKOLPI 780
QY 781 HRLKIDKSFVNDLNEGADTAIIQYVIDLANGLNLETVAEGLIESEAQLORLQKMGCHLGQ 840
DB 781 HRLKIDKSFVNDLNEGADTAIIQYVIDLANGLNLETVAEGLIESEAQLORLQKMGCHLGQ 840
QY 841 GYFLTRPLPAEAMTYLYYPOILDFGPTPLPKYALPETEDAGQVNGDRPLPNSLRE 900
DB 841 GYFLTRPLPAEAMTYLYYPOILDFGPTPLPKYALPETEDAGQVNGDRPLPNSLRE 900
QY 901 NPWTEKLHDYVLKERLOORVKEKLVLIANKIRASININDILYSTVEYRQFNTDRI 960
DB 901 NPWTEKLHDYVLKERLOORVKEKLVLIANKIRASININDILYSTVEYRQFNTDRI 960
QY 961 VLEKENSQMSQGVYTESHNDCRSIINDEIDPCFKGHYLRLYRGRVAVSDLEKADLA 1020
DB 961 VLEKENSQMSQGVYTESHNDCRSIINDEIDPCFKGHYLRLYRGRVAVSDLEKADLA 1020
QY 1021 DCHKELNHYOVKANLVVYVFNENKGLITAEKKTTRYQOEEDLOLMELATQVALAI 1080
DB 1021 DCHKELNHYOVKANLVVYVFNENKGLITAEKKTTRYQOEEDLOLMELATQVALAI 1080
QY 1081 HOGELYEOLLETANIRLQOISSDALTYQGNRYLPFSTJEREMQRLQRIREPALLLCVD 1140
DB 1081 HOGELYEOLLETANIRLQOISSDALTYQGNRYLPFSTJEREMQRLQRIREPALLLCVD 1140
QY 1141 FFKGFNDYGHPRAGDRCLKIADAMAKVAKRPTDLVARYGGEFAIILSETSLGCAINVT 1200
DB 1141 FFKGFNDYGHPRAGDRCLKIADAMAKVAKRPTDLVARYGGEFAIILSETSLGCAINVT 1200
QY 1201 EALOYEVANLAIPIHTVSGTGHVTLISIGIAYVTPERHINPMLVYKADALALYAKAKGRNQ 1260
DB 1201 EALOYEVANLAIPIHTVSGTGHVTLISIGIAYVTPERHINPMLVYKADALALYAKAKGRNQ 1260
QY 1261 WLAYEGSOLPHYDGEV 1276
DB 1261 WLAYEGSOLPHYDGEV 1276
```

RESULT 2  
Y195\_SYNV3  
ID Y195\_SYNV3 STANDARD: PRT: 696 AA.  
AC P74101.  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 79.7 kDa protein s11895.  
GN SLL1895.  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
OX NCBI\_TaxID=1148;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97061201; PubMed=8905231;

RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
RA Miyajima N., Hirosewa M., Sugiyama M., Sasamoto S., Kimura T.,  
RA Hosouchi T., Matsuno A., Muraiki A., Nakazaki N., Naruo K., Okumura S.,  
RA Shimpō S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
RA Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
RT *Synechocystis* sp. strain PCC6803. II. Sequence determination of the  
RT entire genome and assignment of potential protein-coding regions.";  
RL DNA Res. 3:109-136(1996).  
CC -!- SIMILARITY: BELONGS TO THE YEGE/YHDA/YHUK/YJCC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: D90912; BAA18187.1; -  
DR InterPro: IPR001633; DUF2.  
DR InterPro: IPR000160; DUF9.  
DR InterPro: IPR000253; FHA\_domain.  
DR InterPro: IPR000014; PAS.  
DR Pfam: PF00563; FHA\_1.  
DR Pfam: PF00498; FHA\_1.  
DR Pfam: PF00990; GDEF; 1.  
DR SMART: SM00267; DUF1; 1.  
DR SMART: SM00052; DUF2; 1.  
DR SMART: SM00240; FHA; 1.  
DR SMART: SM00091; PAS; 1.  
DR PROSITE: PS50006; FHA\_DOMAIN; 1.  
DR Hypothetical protein; Complete proteome.  
SQ SEQUENCE 696 AA; 79670 MW; 78EB2D4A0EAF6801 CRC64;

Query Match 12.4%; Score 821.5; DB 1; Length 696;  
Best Local Similarity 38.1%; Pred. No. 1.1e-45;  
Matches 171; Conservative 89; Mismatches 184; Indels 5; Gaps 2;

QY 407 AOVASTOLYMAITQOFVRLITQOTAYDPLQPLPMIIFNROLTALLDALYEGKMGVL 466  
DB 239 SIFIRSYFIDITERKVIDQSIIHYQAFYDPLDPLNRFLEKGLVLSNVQNSVLDLV 298  
QY 467 VIANDREKRNESFGHKRGDGLDEVDRLNOKSLPLAAYSPLSRWHDGFTTILTOIS 526  
DB 299 LIGRELOSLNDLGLSHVADVLKITER---LSAHVRLLEDLCRMGDTFILIOQSCR 354  
QY 527 DNOEMIPLCERLSTFOEPFLOGOPYLTASMGISTAPYDGEAESLLKFAETALTRAK 586  
DB 355 NLDELEVVRRLAVLAKRPFIANNPFLTQYAGIACYPNNGDVEILLNRVGIALLNEVK 414  
QY 587 COGNKTYOFTYPODSAPMLDRLTLESIDRLQALNQEVLVYPOVALDTGKLLGEALVR 646  
DB 415 DIGROYCEFYESMNSDLERIOLEHALQALERDEFLLYOPTIIDVSGRLCGEALIR 474  
QY 647 WOHRLGVADVDFIPLAEELGLNHLGOWVLETRACATHOFFRETGRRLMAVNISARQ 706  
DB 475 WOHRLHGVLSGLTIGLETTGLIIPVGEWIMTRAFQHFHMAVADDDFRFIAINLESPQ 534  
QY 707 FODEKWLNVSECLRTKGMPEDELELETESIMMEDIKGVLTALHRLREEVOVAIIDFG 766  
DB 535 FQADDLPLTILRIALAESLPRHRELEETENIVQNTATONLINALQSHGIRLSMDDFG 594  
QY 767 TGYSSLILKOLPIHLRLKIDKSFVNDLNEGADTAIIQYIDLNGNLLETVAIGIESEA 826  
DB 595 TGYSSLILKOLPIHLRLKIDKSFVNDLNEGADTAIIQYIDLNGNLLETVAIGIESEA 826  
QY 827 QLORLQMGCHLGGGYFLTRPLRAEAMMT 855  
DB 655 QARCLYDGLGCKEMOGWFSHPL-SEAEIT 682

RESULT 3  
YAIL\_RHISN  
ID YAIL\_RHISN STANDARD; PRT; 827 AA.  
AC P55552;  
DT 01-NOV-1997 (Rel. 35. Created)  
DT 01-NOV-1997 (Rel. 35. Last sequence update)  
DT 16-OCT-2001 (Rel. 40. Last annotation update)  
DE Hypothetical 91.8 kDa protein YAIL.  
GN YAIL.  
OS Rhizobium sp. (strain NGR234).  
OC Plasmid sym pNGR234.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID=394;  
RN (1)  
RX SEQUENCE FROM N.A.  
RX MEDLINE=97305956; PubMed=9163424;  
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,  
RA Perret X.;  
RT "Molecular basis of symbiosis between Rhizobium and legumes.";  
RL Nature 387:394-401(1997).  
CC -!- SIMILARITY: BELONGS TO THE YEGE/YHDA/YHUK/YJCC FAMILY.  
CC -!- SIMILARITY: CONTAINS 2 PAS (PER-ARN-TSM) DIMERIZATION DOMAINS.  
CC -!- SIMILARITY: CONTAINS 2 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: AE000083; AAB91764.1; -  
DR InterPro: IPR001633; DUF9.  
DR InterPro: IPR000160; DUF9.  
DR InterPro: IPR001610; PAC.  
DR InterPro: IPR000700; PAS-assoc\_C.  
DR InterPro: IPR000014; PAS.  
DR Pfam: PF00563; EAL; 1.  
DR Pfam: PF00990; GDEF; 1.  
DR Pfam: PF00785; PAC; 2.  
DR Pfam: PF00989; PAC; 2.  
DR SMART: SM00267; DUF1; 1.  
DR SMART: SM00052; DUF2; 1.  
DR SMART: SM00086; PAC; 2.  
DR SMART: SM00091; PAS; 3.  
DR PROSITE: PS50113; PAS; 2.  
DR PROSITE: PS50113; PAC; 2.  
DR Hypothetical protein; Repeat; Plasmid.  
KM DOMAIN 12 82 PAS 1.  
FT DOMAIN 212 264 PAC 1.  
FT DOMAIN 265 335 PAC 2.  
FT DOMAIN 338 390 PAC 2.  
SQ SEQUENCE 827 AA; 91818 MW; 0D6CBF1142027FAB CRC64;

Query Match 10.7%; Score 708; DB 1; Length 827;  
Best Local Similarity 28.7%; Pred. No. 3.5e-38;  
Matches 235; Conservative 127; Mismatches 318; Indels 140; Gaps 24;

QY 113 ISPEHNSGHTYVDSCHIQYLLMGVLS-----SLTVPMQDQOLMGINAVHNSKPRRF 167  
DB 62 ILPTAEAD-RYRMD-----IGVLTGEEVSVEPI-----AVPGEVRL 101  
QY 168 TEQEMMALISK-----EVSIAITQSQRVHQOQVQ-----ALVQLETT-- 211  
DB 102 VTR--KSRALITGSSSEKVIYALALDYTECRFAEALQASVEHNSILTEHPVPTAD 159  
QY 212 ----VAQGRPETWOYA-LETVGAVEADGAVLYIAPDLTGSVAQHYQWNLRFDMGNWL 266  
DB 160 PSCEVLEIGPRWEKTYGAPKEALGA-----GWAKAMRPDLGEGYR-----EWAKSL 206

Oy	267	ETSLMOELMRQPSAANEPMMAAAYS--TWKRPRTFSVADLPRTNCPYHGYTL---GELE	321
Db	207	AT-----GEPIDVEFERLLAAAGGYSWRSRRATRA-----EDGSILRLWGYE	250
Oy	322	QRSDWIAPEPESLASENFOSFLIVLADQOV-----GSLILR-----KEKSLYK	367
Db	251	DIDRRKMFELAKSEARFRAIADDAVYMIWVGENGADDDYHSRLMETTGGTAEDQAGK	310
Oy	368	HWAGKRGIDRNILPRISFEAMETQKLVPTWNSSEKLAQVASTQLYMAITQQ-----	421
Db	311	GLMNAVHHDDNNAAVERFYQAFDIREV-----RMEXRLRAGGSAMVIDIQPREASD	365
Oy	422	-----FV-----TRLITQQ-----TAYQPLQPLWMIIFNQTLALDAL	457
Db	366	GTFLGFGVIALDITERNADEBRLLAQKQHMAHNAHNLGRLR-RQFLKEEERLSDDH	424
Oy	458	YEGKMGVLYIAMDREKFRKINESFCHKTGDLLOEVAADRLNOKUSPLAASPLLSRHHG	517
Db	425	APSTRLLATCLDDEGFKAINDAVGRACGDLRLHRTVERLNRFLK---QSDILCRLSGDE	480
Oy	518	FTILLTOISDNOEMIPLCERLLSTFGQEPFFLOGQPIYLTASMSGISTAPYDGETAESLKF	577
Db	461	FVYLVRGVGINSAAEARLLAQGLIDYIENARYELAGHVDQVYVGAASKSDQSLDELKTK	540
Oy	578	AEIALTRAKCGCKMTYQGYFRRQDSAPMLDRLTLESDLRQALTNQEFVLYTPOVALDTCG	637
Db	541	ADIALERAKTGGGGTIVQYEEPKMDADIRARQRMKVSRLHNLAKELFVRYOPLANLRTGQ	600
Oy	638	LLGVATLRMQRHRLRGVARDPVFLPLAEELGLIHHLOWLETFACATNHFREFTRGRLR	697
Db	601	ITTEFLALRMWHPERGOVSPAEPFLAABEJGLIPLESMILROCTAYVM---PPTVS	656
Oy	698	MAVNISAQOFODEKWLNSVLECLARTGMPPEDELELTETESIMMEDIKGTVLLRLHREEG	757
Db	657	VAVNISPLOFRNQIRASIVRNSLSDTGDGLASRLQLEITESVLLSECDSNQITLKEIRQLG	716
Oy	758	VOVAIDDEGTGYSLSILKQLPRLRLKIDKSFVNDLNEGADTAIIOYVIDLANGLNLET	817
Db	717	VIVALDDEGTGYSLSYLRTPPEFKIKVDRSFIDLKRSKESIAIVRAVAAGISGIGIT	776
Oy	818	VAEGISEAQGLORLOKMGCHLGGGCVFLTRPLPAEAMMTYL	857
Db	777	TVEGVERQDDLTITKABGFSDEAGGLTGGCPPLASQAMALL	816
RESULT 4			
ID	YDDU_ECOLI	STANDARD:	PRT: 799 AA.
AC	P616129:	P77708:	P76872:
DT	15-DEC-1998	(Rel. 37,	Created)
DT	15-DEC-1998	(Rel. 37,	Last sequence update)
DT	16-OCT-2001	(Rel. 40,	Last annotation update)
DE	Hypothetical protein yddu.		
GN	YDDU OR B1489.		
OS	Escherichia coli.		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Escherichia.		
OX	NCBI_TaxID=562;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-K12 / MG1655;		
RX	MEDLINE=97426617; PubMed=9278503;		
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Berra N.T., Burland V.,		
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,		
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,		
RA	Mau B., Shao Y.;		
RT	"The complete genome sequence of Escherichia coli K-12.";		
RL	"The complete genome sequence of Escherichia coli K-12.";		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-K12;		
RA	MEDLINE=97251357; PubMed=9097039;		
RA	Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,		

```

RA Itoh T., Nassi H., Kashimoto K., Kimura S., Kitakawa M.,
RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H.,
RA Motomura K., Nakade S., Nakamura Y., Nashimoto H., Nishio Y.,
RA Oshima T., Saito N., Sampei G., Seki Y., Sivasubram S.,
RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA Yamamoto Y., Horikuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996)
CC -1- SIMILARITY: BELONGS TO THE YEGE/YHHA/YHUK/YJJC FAMILY.
CC -1- SIMILARITY: CONSTRAINTS 2 PAS (PER-ARN-T-SIM) DIMERIZATION DOMAINS.
CC -1- SIMILARITY: CONSTRAINTS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isd-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC CC
CC EMBL; AE000246; AAC74562.1; ALT_INIT.
DR EMBL; D90789; BAA15144.1; -.
DR EMBL; D90790; BAA15154.1; -.
DR EMBL; D90791; BAA15160.1; -.
DR EcoGene; EG13792; ycdU.
DR InterPro; IPRO01633; DUF2.
DR InterPro; IPRO00160; DUF9.
DR InterPro; IPRO01610; PAC.
DR InterPro; IPRO00700; PAS-assoc_C.
DR InterPro; IPRO00014; PAS.
DR Pfam; PF00563; EAL; 1.
DR Pfam; PF00990; GDDEF; 1.
DR Pfam; PF00989; PAS; 2.
DR SMART; SM00267; DUFL; 1.
DR SMART; SM00052; DUF2; 1.
DR SMART; SM00066; PAC; 2.
DR SMART; SM00091; PAS; 2.
DR PROSITE; PS00112; PAS; 2.
DR PROSITE; PS00113; PAC; 1.
KW Hypothetical protein; Repeat; Complete proteome.
FT DOMAIN 10 79 PAS 1.
FT FT 134 205 PAS 2.
FT FT 208 260 PAC.
SQ SEQUENCE 799 AA; 90260 MW; 4CE770DD2A6B39D5F CXC64;

Query Match 10.3%; Score 684.5; DB 1; Length 799;
Best Local Similarity 33.3%; Pred.No.1.le-36;
Matches 172; Conservative 99; Mismatches 204; Indels 41; Gaps 12;

QY 355 SLILLRKEKSLVKKHNA-GKRGIIDRNILPRLSFEAMEETOK-----LYPTWNNSE 403
| | | | | | | | | | : : : : : : : : : : : : : : : : : : : : : : :
DB 292 SFFALRNGMPI--HNVASSHGAELIONA-----QSMSATIRORDAGAPGIIQIKSSGE 343
| | | | | | | | | | : : : : : : : : : : : : : : : : : : : : : : :
QY 404 -----RKLAQVASTQLUYAMITQQFYRLTLTOOTAVDPPLTPLOPNMTIFNRQTLALDDALY 458
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 344 TSAFIERRVADISQHMAALALEBEKSRQHEIOLIQDPMPTGG-----KRNNTAHNLDDLIV 398
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 459 ECKRWG-VLVIAMDFEKRINESFGHKTDGLLOEVADRNLNOKISPLAAYSPLSRMHGD 516
| | | | | | | | | | : : : : : : : : : : : : : : : : : : : : : : :
DB 399 D-KAVSPVYLLIGVDHIODVIDSLGYAMADALLEVVNRFREKLKP----DQYLCRIGCT 453
| | | | | | | | | | : : : : : : : : : : : : : : : : : : : : : : :
QY 517 GTTILLTQSDNQEMIPICERLTSTFOEFPILOGOPIVTASMGSTAPYD-GETAESLL 575
| | | | | | | | | | : : : : : : : : : : : : : : : : : : : : : : :
DB 454 QF-VLVSLENDYSNTQTQADELRNVASKIMIDDKPFPLTSTGIS--YDLGKNRDYLL 509
| | | | | | | | | | : : : : : : : : : : : : : : : : : : : : : : :
QY 576 KAFAIALTRAKOCGNKYTFYRPDSAPMLDLTLESIDLQAALTQNEFVLFPQVALDT 635
| | | | | | | | | | : : : : : : : : : : : : : : : : : : : : : : :
DB 510 STAHNMADIRKNGNGMQFSPANNENMYKEELVLGAALKKAISNNOLKTVYQPQIFAET 569
| | | | | | | | | | : : : : : : : : : : : : : : : : : : : : : : :
QY 636 GLTGVREALVRMQRPLRGADVFIPLAEELGLINHGQVYLETRACATHQHFRETGR 695
| | | | | | | | | | : : : : : : : : : : : : : : : : : : : : : : :

```

```

DB 570 GELTGIEALARNHDP.LGHVPPSR.FIPLAEIIEIGENIGMWIYAECROLAENRSNIHT 629
QY 696 LMAVNIASAROFODEKWLNSVLECLARTGMPDELEITESLAMEDIKGTVLLHRLRE 755
DB 630 PALSVLNSALHFRSNQLPNOVSDAMHAMGIDGHLVTEIESMMMEHDPEIFKRIQLRD 669
QY 756 BEVOVAIDDEGCYSSLSLTKOLPHRLKIDKSFVNDLNEGADTAIOYVIDLANGLNL 815
DB 690 MGVSLSVDFGFGFSGLSRLVS.LPVEIKIDKSFVDRCLERKRLALLLEAITSIGOSLNL 749
QY 816 ETVAEIESEAO.LQRLQKMGCHLGOGYFLTRPPLPAE 851
DB 750 FYVAEVEERKEQEFEMLRKIHCRVIOGYFSPRLPAE 765

RESULT 5
YD54_MYCTU STANDARD: PRT: 623 AA.
AC 011024:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 67.7 kDa protein RV1354c.
GN RV1354C OR MT1397 OR MTCY02B10.18C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV:
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Horsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN-CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Kouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
CC -! SIMILARITY: BELONGS TO THE YEGE/YHDA/YHKK/YJCC FAMILY.
CC -! SIMILARITY: SOME TO SYNECHOCYSTIS SP. SLR0359.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z75555; CAA99987.1; -
CC EMBL: AE007012; AAK45660.1; -
CC HSSP: P03622; ZIFO.
CC TIGR: MT1397; -
CC Tuberculist: RV1354c; -
CC InterPro: IPR001633; DUF2.
CC InterPro: IPR000160; DUF9.
CC InterPro: IPR003018; GAF.

```

```

DR Pfam; PF00563; EAL; 1.
DR Pfam; PF01590; GAF; 1.
DR Pfam; PF00990; GGEF; 1.
DR SMART; SM00267; DUF2; 1.
DR SMART; SM00052; DUF2; 1.
DR SMART; SM00065; GAF; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 623 AA; 67653 MW; ED54E5FA97191B7 CRC64;

Query Match 9.4%; Score 621; DB 1; Length 623;
Best Local Similarity 32.5%; Pred. No. 1e-32;
Matches 160; Conservative 85; Mismatches 212; Indels 36; Gaps 9;

QY 399 WNRSEKRLAOVAST---QLYMAITQOFVTRLIHQGRVDPDLPMMIIFNRQTLTALD 455
DB 152 WHEAEIENALMTITLFAQVQARVAERARLYLD---HDLTGL-----HNRALLQHD 203
QY 456 -ALYEGK--MVGVLVYAMDREKRINESFGHRTGDGLQEVADRNLKSLPLAAYSPLLSR 512
DB 204 QRLAPQPGPVALFLDIDRLKAINDYLGHAAGDQFIHVAQRIGDAL-----VGESLIAR 259
QY 513 WHDGDFTILLTQISDNQEMIPLEBRLLSTQPEFFLQGPITYITASGISATAYDGETAE 572
DB 260 LGSDEFVLLIPASPSADAAQPIAERLDQKDHVAIGGEVLTFTVTSIGVASGTPGQHTPS 319
QY 573 SLTKFAEIALTRAKCOGKNTYQFRRPQDSAPMLDRLTLESDDLROALTNQEFVLYPOPOVA 632
DB 320 DILRRADQALAAKHNAGGSVALFTADMSVSGSLRNDIELHLRGLIESDRLVLYPEVD 379
QY 633 LDTGKLGYEALVRQWHPRLGQVAPVFIPLAEGLINHLGQVLETCATQHFFRE- 691
DB 380 LRTGDIIVGTVALVRWQHPRLGLAPGCFIPVASINLAGELDRWVLRAC---NEFSEW 435
QY 692 ----TGRRLMAVNIASAROFODEKWLNSVLECLARTGMPDELEITESLAMEDIKGY 747
DB 436 QSAGLCHDALRINVASAGLVYGGFVPAVDITGQHLDASSVCLTEVNVVQDLHTAR 495
QY 748 VLLHRLREGOVAYIDFGTGYSSLSLTKOLPHRLKIDKSFVNDLNEGADTAIOYVY 807
DB 496 ATLARIKEVGHVHAIIDDFGYSAISLGLTLPIDTLKIDKTFVRLQGTNMSDLYIRGIM 555
QY 808 DLANGINLETVAIGTSEAO.LQRLQKMGCHLGOGYFLTRPPLPAEMTYIYVQIILDFG 867
DB 556 TLAEGFQLDVVAEGVETEAARILLDQRCYRAGGFLESRPVEAMRHML-----S 606
QY 868 TPPLKVALPETE 880
DB 607 ARRLPETCIPARD 619

RESULT 6
YCIIR_ECOLI
ID YCIIR_ECOLI STANDARD: PRT: 661 AA.
AC P77334:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yciir.
GN YCIIR OR B1285.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";

```

```

RL Science 277:1453-1474(1997).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN-K12:
RX MEDLINE-97251357; PubMed-9097039:
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itch T.,
RA Kasai H., Kasahimto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Samped G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA Takeuchi K., Takeuchi Y., Mada C., Yamamoto Y., Horuchi T.,
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [3]
RN SEQUENCE OF 1-557 FROM N.A.
RA Rita Z.R., Arralano C.A.;
RL Submitted (FEB-1996) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE YESE/YHDA/YHUK/YUCC FAMILY.
CC -1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
CC -1- CAUTION: REF.3 SEQUENCE DIFFERS EXTENSIVELY TO THAT SHOWN HERE DUE
CC TO MANY FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000226; AAC74367.1; -
CC DR EMBL; D90767; BAA14847.1; -
CC DR EMBL; D90766; BAA14839.1; -
CC DR EMBL; L40788; -; NOT ANNOTATED_CDS.
CC EcceGene; EG13415; yjcR.
CC DR InterPro; IPR001633; DUF2.
CC DR InterPro; IPR000160; DUF9.
CC DR InterPro; IPR000014; PAS.
CC DR Pfam; PF00563; EAL; 1.
CC DR Pfam; PF00990; GGDEF; 1.
CC DR Pfam; PF00989; PAS; 1.
CC DR SMART; SM00267; DUF1; 1.
CC DR SMART; SM00052; DUF2; 1.
CC DR SMART; SM00091; PAS; 1.
CC DR PROSITE; PS00112; PAS; 1.
CC KW Hypothetical protein: Complete proteome.
CC DOMAIN 109 179 PAS.
CC FT CONFLICT 4 4 V-> L (IN REF. 3).
CC FT CONFLICT 26 27 SS-> AA (IN REF. 3).
CC FT CONFLICT 113 113 A-> P (IN REF. 3).
CC FT CONFLICT 175 175 V-> L (IN REF. 3).
CC FT CONFLICT 256 256 D-> H (IN REF. 3).
CC FT CONFLICT 344 344 P-> G (IN REF. 3).
CC FT SEQUENCE 661 AA; 74664 MW; 991D47CDE46AF01 CRC64;
SQ
Query Match 9.2%; Score 607; DB 1; Length 661;
Best Local Similarity 31.6%; Pred. No. 9e-32;
Matches 144; Conservative 110; Mismatches 165; Indels 36; Gaps 8;
OY 401 RSEKLAOVASTQLYMAITQOFVTRLLTQOTAYDPLTQLPNNIIFRQLTALLDALY-- 458
DB 228 RAOERLRILANT-----DSITGLP-----NRNMODLIDHAINH 261
OY 459 -EGKAVGVIVMDPRKRNESGKHTGDLLOEVDARNOKLSPLAASPLSRHGG 517
DB 263 ANNNKVVGVYLDLDNRKKVNDAYGHLFGDQLRDVSLAI--LSCV-EHDQVYLRGGE 317
OY 518 FTLLTQISDNQEMIPICRLSTFQEPFLQOGPIYLTASMGISTAPYDGETAESLKF 577
DB 318 FLVLAISNTSQS-ALFAMASRIILRLRPRRIGLIEVYTSQSVGIALSPRHGSSTAIITH 376

```

```

OY 578 AEIALTRACOGKNYQFRRPODSAPMLDRLLLESRLRALNNOEVLKFOPOVALDNCK 637
DB 377 ADTAMTAKGGRGOCFCVTPENQOFVEYLMVDNLRLRALENDOLVHYQPITW-RGE 435
OY 638 LIGVEALVKNOPRRLGQVDPVEIFPLAEELGLINHLGQVWLETCATQHOFRENGRRLR 697
DB 436 VASLEALVWQSPERGLIPPLDPIISAESGLVPLGRVIT-LDVYRQAKRDKGINLR 494
OY 698 MAVNISARQFODEKWLNSVLECLKRTGMPDELELETESLMEDIKGTVLLHRLREB 757
DB 495 VAVNISARQADQITFTALKQVLELNFECPLDVELTESCLLENDELALSVYQFSQUG 554
OY 758 VOYADDFETGYSLSLTKQLPIHRKIDKSYNDLINCAGTATIQYVIDLANGINLET 817
DB 555 AQVHLDDEFTGTGYSLSQLARFPIDAKLDQVPRDTHKQVQSQSLVRAIVAAQALNLQY 614
OY 818 VAEIGSEAOLORLQKMGCHLGQGYFLTRPDAEA 852
DB 615 IAEVSEAKEDAFLLKNGINERQGLFKAQMPRAVA 649
RESULT 7
YHUK_ECOLI
ID YHUK_ECOLI STANDARD: PRT: 651 AA.
AC P37649;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein yhjK.
GN YHUK OR B3529.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-94316500; PubMed-8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes.";
RL Nucleic Acids Res. 22:2576-2586(1994).
RN [2]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE-9420866; PubMed-10493123;
RA Fountoulakis M., Takacs M.-F., Berndt P., Langen H., Takacs B.;
RT "Enrichment of low abundance proteins of Escherichia coli by
RT hydroxyapatite chromatography.";
RL Electrophoresis 20:2181-2195(1999).
CC -1- SIMILARITY: BELONGS TO THE YESE/YHDA/YHUK/YUCC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U00039; AAB18506.1; -
CC DR EMBL; AE000429; AAC76554.1; -
CC EcceGene; EG12256; yhjK.
CC DR InterPro; IPR001633; DUF2.
CC DR InterPro; IPR000160; DUF9.
CC DR InterPro; IPR003660; HAMF.
CC DR Pfam; PF00990; GGDEF; 1.
CC DR Pfam; PF00990; GGDEF; 1.
CC DR Pfam; PF00672; HAMF; 1.
CC DR SMART; SM00267; DUF1; 1.
CC DR SMART; SM00052; DUF2; 1.
CC KW Complete proteome.
SQ SEQUENCE 651 AA; 73080 MW; 9C914999411D9E8D CRC64;

```

```

CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN
CC this SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/or\_send\_an\_email\_to\_license@isb-sib.ch).
CC -----
CC EMBL, X63841; CAA45329.1;
CC PIR, S18623; S18623.
CC HSSP, P23222; IDRM.
CC InterPro: IPR001633; DUF2.
CC InterPro: IPR000160; DUF9.
CC InterPro: IPR003018; GAF.
CC InterPro: IPR001610; PAC.
CC InterPro: IPR00700; PAS-assoc_C.
CC InterPro: IPR000014; PAS.
CC Pfam: PF00563; EAL; 1.
CC Pfam: PF01590; GAF; 1.
CC Pfam: PF00990; GGDGF; 1.
CC Pfam: PF00989; PAS; 1.
CC SMART: SM00267; DUF1; 1.
CC SMART: SM00052; DUF2; 1.
CC SMART: SM00065; GAF; 1.
CC SMART: SM00086; PAC; 1.
CC SMART: SM00091; PAS; 1.
CC PROSITE: PS01112; PAS; 1.
CC PROSITE: PS01113; PAC; 1.
CC Hypothetical protein.
CC DOMAIN 185 255 PAS.
CC FT DOMAIN 263 313 PAC.
CC SMOQUENCE 735 AA; 80476 MW; CFF94DDE06440B9A CRC64;

Query Match 8.1%; Score 535; DB 1; Length 735;
Best Local Similarity 24.7%; Pred. No 5.2e-27;
Matches 221; Conservative 125; Mismatches 338; Indels 212; Gaps 28

QY 9 DELNVNKKFRALTLRETLQVLYEARIFLGVDRVYIY-PSDSGSEVLAANVRAL 67
DB 32 DLARINDTFH-----LQRYSLFGVHNEGRG-IAACVIDWR 69
QY 68 PSLIGLHP--VEDIPQARELGNQKMIADVANHRRKSHESGISPTENSNG---- 121
DB 70 P---GLAFPAETGESHPLTAAGDP---LIAQAAARRRKEITIGR---TRDLTGVLG 120
QY 122 ---HTTVDSCHIGYLLAMGVLSLTVPVKMDOLMGIMAVHSHKPRRFEQEMETALL 178
DB 121 PFSH-----GVTFLETPVNVHGRWG---HFCVDPDAEHMTAERQ 162
QY 179 SKVSLAITSQSLSFQVHQOQVQYQEAIVQRLETTVAQVGDREPTQYALFVGAVADGA 238
DB 163 AFKCIAAVLACILARRSGTEGLVSEARRAM----- 192
QY 239 VLIYAPDLTGSVAQYQWNILRFDMQNMLETSLIMELMNGQESAAMEPAAVQSTWEKRP 298
DB 193 -----LDTSIDAIVADEGALVEFHNAAEALFGHTR 225
QY 299 ETSVAELPPTCVPRGYTLGELQSDMTAPESISAENFOSFLVPLPAADQWVGSLL 358
DB 226 GVGGRPMET-ILPAHYIDRHQGGEMRHATGENHIMRLVE--VEALRAD---GS--V 276
QY 359 LRKESLYVHMGKRGIDR---RNILRSLSEAWETQRYLPTNRSERKIAQVASTOLY 415
DB 277 EPALVELVNEHRAGKRLFSAPVRDISDR-----SRRLERLAFDMH 320
QY 416 MATTQGF-VTRLTQQTAYDPLTQLPNIILFRQLTALLDALYEGKMGVAVIADRRK 474
DB 321 TGLSNRTGLRLCTGRPT-----RSGGAVVLMRLDG 352
QY 475 RINSEGCHKGGDLLEVDNRLNQKLSPLAAYSPLLSRWNGDFTLLTQISDNQEMIP 534

```



```

DB 353 YVKTSGDDMAEPMIVETANLRLMLPQEC---LGRGESEFEVTVWQPCAAE---L 405
      : ||| : : : : : : : : : : : : : : : : : : : : : : : : :
QY 535 CERLLTFOEPFLOGOPYLTASMGISTAPYDGETAESLLKFAETALTRACOGKNYQ 594
      : : : : : : : : : : : : : : : : : : : : : : : : :
DB 406 AETLTIORLASAIESGRRRYLRGLGVERPGD---ATYLLRDAEWA---ARDCROGLLH 460
      : : : : : : : : : : : : : : : : : : : : : : : : :
QY 595 FPRPDASAPMLDRLTLESRLQALTNQ---EFVLYFOPQVALDTGKLLGVEALVRMHPRL 652
      : : : : : : : : : : : : : : : : : : : : : : : : :
DB 461 FAEHM-RACHOOBLEMLEMLRDVIQRTTSALSHYQPVVSARTGLVGEFALRWYSETH 519
      : : : : : : : : : : : : : : : : : : : : : : : : :
QY 653 GOVAPVFTPLAEELGLINHLGQWLET---ACATQHFEFRETG-PRLLMAVNISAROFQ 708
      : : : : : : : : : : : : : : : : : : : : : : : : :
DB 520 GVPSPALFPLAAGFAEELGAWVETASACAGNNVRRRANGLAPMHAIALNSTEVV 579
      : : : : : : : : : : : : : : : : : : : : : : : : :
QY 709 DEKMNSVLECLKRTGMPPEDELEETESLAMEDIKGTAVLLHRLREGVQVADIDFGTG 768
      : : : : : : : : : : : : : : : : : : : : : : : : :
DB 580 APDLIERVROTMAFHGLPQCVCFELETESAILNQPEIAIETLSRLALGCTTADIDFGTG 639
      : : : : : : : : : : : : : : : : : : : : : : : : :
QY 769 YSSLSTLKOLPIHRLKIDKSFVNDLNEGADTAIIQVVIDLANGLNLVETVAEGISEAOL 828
      : : : : : : : : : : : : : : : : : : : : : : : : :
DB 640 YSSLSTLQRLPMQVYLKIDKSFVLDVNDNSRSEIVRMIMANGLSVYAEGETTGAL 699
      : : : : : : : : : : : : : : : : : : : : : : : : :
QY 829 QRLQKMGCHLGGYFLTRPLPAEAMTYLYYQIILDFGPTPLPKVALPETETEAG 884
      : : : : : : : : : : : : : : : : : : : : : : : : :
DB 700 QILRQMGCDRAQGFELFGRAHPGVAGT-----LPETLAPTG 735
      : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT_ 9
YJCC_ECOLI
ID YJCC_ECOLI STANDARD; PRT; 528 AA.
AC P32701; P76789;
DT 01-OCT-1993 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein yjcc.
OS YJCC OR B4061.
OC Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MG1655;
RX MEDLINE=94089392; PubMed=8265357;
RA Blattner F.R., Burland V.D., Plunkett G. III, Sofia H.J.,
RA Daniels D.L.,
RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the
RT region from 89.2 to 92.8 minutes."
RL Nucleic Acids Res. 21:5408-5417(1993).
CC -1- SIMILARITY: BELONGS TO THE YEGE/YHDA/YHJK/YJCC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U00006; AAC3155.1; -
CC EMBL; AE000479; AAC77031.1; -
CC DR Ecogène; EG11938; yjcc.
CC DR InterPro; IPR001633; DUF2.
CC DR Pfam; PF00563; EAT_1.
CC DR SMART; SM00052; DUF2; 1.
CC DR Hypothetical protein; Complete proteome.
CC KW SEQUENCE 528 AA; 60801 MW; F6E4819954912F31 CRC64;
SQ

```

```

QY 608 LTLESRLQALINQFVLYFOPQVALDTGKLLGVEALVRMHPRLGQVA-PDVEFIPLAE 666
      : : : : : : : : : : : : : : : : : : : : : : : : :
DB 269 LSPKRLQALAEKHOLCLHYDPIIDIKTEKICGALLRMGEO-GQIMNAEPIPLAEK 327
      : : : : : : : : : : : : : : : : : : : : : : : : :
QY 667 LGLINHLGQWLETACATQHFEFRETGRRLR-----MAVNISARQODEKWLNSVLECL 720
      : : : : : : : : : : : : : : : : : : : : : : : : :
DB 328 EGMIGQITDYVID-----NVFRDLGDLATHADRVYSINLSASDFHTSRILARINOKT 380
      : : : : : : : : : : : : : : : : : : : : : : : : :
QY 721 KRTGMPPEDELEETESLAMEDIKGTAVLLHRLREGVQVADIDFGTSSSLTKOLPI 780
      : : : : : : : : : : : : : : : : : : : : : : : : :
DB 381 EQAVARPDQIKVEYEHAFELVDKMTPIIL-AFRQAGYVAIDDGIGTSSNHLNLSLVN 439
      : : : : : : : : : : : : : : : : : : : : : : : : :
QY 781 HLKIDKSFVNDLNEGADTAIIQVVIDLANGLNLVETVAEGISEAOLQRLQKMGCHLQ 840
      : : : : : : : : : : : : : : : : : : : : : : : : :
DB 440 DLKIDKSFVETLTTHKNSHLAEHIELASHLGKLTAEVETEEQVNWMLKRGVRYCQ 499
      : : : : : : : : : : : : : : : : : : : : : : : : :
QY 841 GFELTRPLPAEAMTYLYYQIILDFGPTPLPKVALPETETEAG 885
      : : : : : : : : : : : : : : : : : : : : : : : : :
DB 500 GMEFAKAMPQVFMQWM-----EQLPARELTRGQ 528
      : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT_ 10
YJLE_ECOLI
ID YJLE_ECOLI STANDARD; PRT; 782 AA.
AC P75800;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yjle.
OS Escherichia coli.
OC Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.,
RT "The complete genome sequence of Escherichia coli K-12."
RT Science 277:1453-1474(1997).
RL [2]
RP SEQUENCE FROM N.A.
RX STRAIN-K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horinouchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map."
RT DNA Res. 3:137-155(1996).
CC -1- SIMILARITY: CONTAINS 1 DUF2 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000185; AAC73920.1; -
CC DR EMBL; D90721; BAA35528.1; -
CC DR EMBL; D90722; BAA35536.1; -
CC DR Ecogène; EG13476; yjle.
CC DR InterPro; IPR001633; DUF2.

```

Query Match 6.0%; Score 400; DB 1; Length 528;  
 Best Local Similarity 33.0%; Pred. No. 1.8e-18;  
 Matches 94; Conservative 60; Mismatches 99; Indels 32; Gaps 6;





Db 587 PTVSQARPFVEVSOLLAKIOIEMQOLIFEVENSALTNTVKAQITLOHDELGCQIAD 646  
 Oy 764 DEFGYSSLSLTKOLPIHRLKIDKSFVNDLNEGADTAIIQYVIDLANGLNETVAEGIE 823  
 Db 647 DFGYASVARKVNNADLTKIDGSEFINVSNLIDYOIVASICHLRKKMLVVAEVE 706  
 Oy 824 SEAOIQRLOKMGCHLGQGYFLTRPLP 849  
 Db 707 NEIREAVLSIGIDYMGQYLGKPP 732

RESULT 12  
 YD57\_MYCTU STANDARD; PRT; 307 AA.  
 ID YD57\_MYCTU  
 AC 011027;  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 33.9 kDa protein RV1357c.  
 GN RV1357C OR MT1400 OR MTCY02B10.21C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 NX MG1\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cline S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.,  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RA "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains." to the EMBL/GenBank/DBJ databases.  
 RL Submitted (AFR-2001) to the YEST/YHDA/YHCK/YJCC FAMILY.  
 CC -1- SIMILARITY: BELONGS TO THE YEST/YHDA/YHCK/YJCC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial/  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: Z75555; CA99962.1; -;  
 DR EMBL: AE007012; AAK45663.1; -;  
 DR TIGR: MT1400; -;  
 DR Tuberculist: RV1357c; -;  
 DR InterPro: IPR001633; DUF2.  
 DR Pfam: PF00563; EAL; 1.  
 DR SMART: SMO0052; DUF2; 1.  
 KW Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 158 178 POTENTIAL.  
 FT TRANSSEM 203 223  
 KW Hypothetical protein; Transmembrane; Complete proteome.  
 FT SEQUENCE 307 AA; 33982 MW; 3DC597A6AB3BAC48 CRC64;

Query Match 5.6%; Score 368.5; DB 1; Length 307;  
 Best Local Similarity 35.9%; Pred. No. 8.5e-17;  
 Matches 89; Conservative 46; Mismatches 106; Indels 7; Gaps 3;

Oy 610 LESDRQALTNQEPVLYPOVALDTGKILGYEALVRMOPHPLGOVAPVFIPLAEIQL 669  
 Db 57 LSTSRVALDRGEFFLVYOPRIRLADNRITGEALLRHEHPTLGTLLPRFIDRAENGL 116  
 Oy 670 INHLGQWLETCATGHOHFFRE--TGRRLRMVNISARQFOEKVNSVLECKTKRTPMP 727  
 Db 117 MYPLTAFVLEQAC-RHVSWMRSHSTDPQFVSVNVSASTICDPGLVVEGLGETGLPA 175  
 Oy 728 EDLELEITESLMEDIKGTVLLRLREGVQVAIDDEFGYSSLSLKQLPPIHRLKIDK 787  
 Db 176 HALQLELAEDARLSDEKRAVTRLQELSLAGVGAIDDEFGIGSSLAIVLPFLVDVVKLG 235  
 Oy 768 SFV----NDLNEGADTAIIQYVIDLANGLNETVAEGIESAOLQRLKMGCHLGQGYE 843  
 Db 236 KTEICLDGDIQRLANEOTTRAMIDLGKLGITVTRAKVETPSQARLARAFCKAAQGMH 295  
 Oy 844 ITRPLPAE 851  
 Db 296 FAKALPVD 303

RESULT 13  
 YLAB\_ECOLI STANDARD; PRT; 516 AA.  
 ID YLAB\_ECOLI  
 AC P77473;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein ylab.  
 GN YLAB OR B0457.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 NX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RA "The complete genome sequence of Escherichia coli K-12."  
 RL Science 277:1453-1474(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,  
 RA Duncan M., Federapfel N., Hyman R., Kalman S., Komp C., Kurdi O.,  
 RA Lew H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.;  
 RA Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE YEST/YHDA/YHCK/YJCC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial/  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: AE000152; AAC73559.1; ALT\_INIT.  
 DR EMBL: U82664; AAB40212.1; ALT\_INIT.  
 DR Ecocore: EG14237; ylab.  
 DR InterPro: IPR001633; DUF2.  
 DR Pfam: PF00563; EAL; 1.  
 DR SMART: SMO0052; DUF2; 1.  
 KW Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 6 26 POTENTIAL.  
 FT TRANSSEM 242 262





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2002, 18:53:41 ; Search time 60.43 Seconds

(without alignments)  
3652.847 Million cell updates/sec

Title: US-09-272-809-2

Perfect score: 6614

Sequence: 1 NMPNRSLEDFLRNVIKFRH.....GRNQWLAYEGSQLPHVDEY 1276

Scoring table: BLOSUM62

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database : SPTREMBL\_19:\*

1: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	922	13.9	1415	16	Q9HV18
2	882	13.3	748	16	Q9A3V7
3	865	13.1	1244	16	Q55576
4	862.5	13.0	1245	16	Q915W1
5	858.5	13.0	772	16	Q92KJ3
6	830.5	12.6	760	16	Q916K5
7	811.5	12.3	783	16	Q9HYT3
8	811	12.3	685	16	Q91310
9	810.5	12.3	899	16	Q9HUF2
10	809	12.2	687	16	Q9HW35
11	801.5	12.1	840	16	P72843
12	797	12.1	1578	16	P74400
13	789.5	11.9	892	2	Q9LAF1
14	785	11.9	724	16	Q984Y8
15	780.5	11.8	653	16	Q981Y4
16	774.5	11.7	808	16	Q98G38

17	769.5	11.6	614	16	Q98J06	Q98J06 rhizobium 1
18	768	11.5	951	16	Q9HTC4	Q9HTC4 pseudomonas
19	758.5	11.5	774	2	Q9AF11	Q9AF11 vibrio para
20	757	11.4	865	16	Q9K065	Q9K065 vibrio chol
21	752.5	11.4	856	16	Q9RS28	Q9RS28 delinococcus
22	746	11.3	864	16	Q91243	Q91243 pseudomonas
23	739.5	11.2	732	16	P73380	P73380 pseudomonas
24	734.5	11.1	635	2	Q935Y8	Q935Y8 synechocyst
25	731.5	11.1	634	2	Q93J80	Q93J80 streptomyces
26	731.5	11.1	842	16	Q9A666	Q9A666 caulobacter
27	717.5	10.8	564	16	Q92WP4	Q92WP4 rhizobium m
28	714	10.8	742	16	Q930X2	Q930X2 rhizobium m
29	714	10.8	1051	16	Q92YNO	Q92YNO rhizobium m
30	705.5	10.7	805	16	Q9RX49	Q9RX49 delinococcus
31	703.5	10.6	749	16	Q943J1	Q943J1 bacillus su
32	700.5	10.6	809	16	Q9ABX9	Q9ABX9 caulobacter
33	699.5	10.6	746	2	Q92584	Q92584 streptomyces
34	689	10.4	752	2	Q87376	Q87376 acetobacter
35	687	10.4	543	16	Q9K8N4	Q9K8N4 bacillus ha
36	686	10.4	853	16	P72746	P72746 synechocyst
37	678	10.3	735	16	Q9A9V4	Q9A9V4 caulobacter
38	677.5	10.2	772	16	Q97L01	Q97L01 clostridium
39	672	10.2	558	16	Q9HTQ9	Q9HTQ9 pseudomonas
40	671.5	10.2	446	16	Q983J9	Q983J9 rhizobium 1
41	670	10.1	714	2	Q9RCY2	Q9RCY2 streptomyces
42	665.5	10.1	738	16	Q98BE2	Q98BE2 rhizobium 1
43	663.5	10.0	554	16	Q985V5	Q985V5 rhizobium 1
44	663.5	10.0	814	16	Q67220	Q67220 aquifex aeo
45	661	10.0	740	2	Q87378	Q87378 acetobacter

## ALIGNMENTS

RESULT	ID	Q9HV18	PRELIMINARY:	PRT: 1415 AA.
AC	Q9HV18	01-MAR-2001 (TREMBLrel. 16, Created)		
DT	01-MAR-2001	(TREMBLrel. 16, Last sequence update)		
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)		
DE	HYPOTHETICAL PROTEIN PA4601.			
GN	PA4601.			
OS	Pseudomonas aeruginosa.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;			
CC	Pseudomonas.			
OX	NCBI_TaxID=287;			
RN	11			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-ATCC 15692 / PA01;			
RX	MEDLINE-20437337; PubMed-10984043;			
RA	Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,			
RA	Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,			
RA	Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,			
RA	Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,			
RA	Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,			
RA	Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;			
RT	"Complete genome sequence of Pseudomonas aeruginosa PA01, an			
RT	opportunistic pathogen.";			
RL	Nature 406:959-964(2000).			
DR	EMBL; AE004874; AAC07989.1; -			
DR	InterPro; IPR001633; DUF2.			
DR	InterPro; IPR00160; DUF9.			
DR	InterPro; IPR003015; HLH_Myc.			
DR	InterPro; IPR001610; PAC.			
DR	InterPro; IPR000014; PAS.			
DR	InterPro; IPR000700; PAS-assoc_C.			
DR	Pfam; PF00990; GCDERF.1.			
DR	Pfam; PF00785; PAC.3.			
DR	SMART; SM00267; DUF1.1.			
DR	SMART; SM00052; DUF2.1.			
DR	SMART; SM00086; PAC.3.			
DR	SMART; SM00091; PAS.4.			

DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; UNKNOWN\_1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 1415 AA; 159663 MW; F12347DD0BA89E2 CRC64;

Query Match 13.9%; Score 922; DB 16; Length 1415;  
 Best Local Similarity 34.6%; Pred. No. 4.3e-53;  
 Matches 250; Conservative 112; Mismatches 263; Indels 98; Gaps 19;

OY 171 EMEYALLSKREVSIAITQSLSR-OVHOQOQOEALVORLETTVAOYGDPEETWOYALETV 229  
 DB 755 EW---ALINGLHOTATNRQLRLGLRLRVNHAIGDRK--LAELREMPGHLFALD-- 807  
 OY 230 GOAVEADAVLYIADLTGSAQHYNQVWMLRFDGMWLETSLMOELMR-----GQ 278  
 DB 808 --CLNADKRKIP-----ELRI-----VLMDEHDFRGLGIARDISQ 844  
 OY 279 PSAMEPMAAVOSTWEKRPFTSVAPLPPYNCVPHGY--TLGELQSRSDWAP----- 329  
 DB 845 QRRARRELMAATVEH---STAIAIVTD--PAGYIVQVNDSPSRITGYPADVLDOQ 897  
 OY 330 PESTLSAE---NFOSELIIVPLAADQWVGSITLLREKESLYVMAGKNGI-DRNIIPLRL 384  
 DB 898 PRLTADQOANQLKLVIALSHSGSWGEIILQKRKTGELIYVWGTAVRDESGDL-V 955  
 OY 365 SEAMEETOKLVPTWNRSEKRLAQVASTOLYMAITQOEVTRILITQOTAYDPLTQLPNMTI 444  
 DB 956 SFVCFES-----DISERKASE-----RIHRLAYYDALTHLPNRL 991  
 OY 445 ENKQITLLALLDALYEGKAVGVLYVIMDRFRKINESFGHKTGGGLLOEVADELNOKLSPLA 504  
 DB 992 FODRLHTLQAQERNQWVLMFLDLDRPKINDSLGNAAGRMLOEVATR---LSACV 1047  
 OY 505 AYSPLLSRMHGDGFTILLTQISDNQ---EMIPLCERLSTFOEPFFQOGPIYLTASMG 560  
 DB 1048 SODTVAMAGDEFTLLPSCGDRRIALKRAIOVAEILIGRLARFTELEGREFYTAGISG 1107  
 OY 561 ISTADYGETAESLKFELIALTRAKCOGKNTYOFYRPODSAPMLDRLTLESDLRQALTN 620  
 DB 1108 VALSPQDAEELSLMKKNADTAMYNHAKEMKKNFQFOAEMNARALELESDLRRALEL 1167  
 OY 621 QEFVYPOPYALDTGKLLGVEALVRWQHPRLGOVAPVPTPLAEELGINLNGWVLET 680  
 DB 1168 GEFVLHPOPTGGDRRLTGAELRKLWQHPRLGVPSEFIVLEIGLVAOVGMVLAE 1227  
 OY 661 ACATQHFFRETGRRLRAVNVISARQFODEKVLNVLKELKRTGAPPEDELEITESLMM 740  
 DB 1228 ACKOLRSWHKAKVRYKVSVNLISARQFADGQGERIAIIVETGIRPACLELEITESLIM 1287  
 OY 741 EDIKGTVVLHRLREGEYQVAIDFGTGYSSLSILKQLPIHRLKIDKSFVNDLNEGADT 800  
 DB 1288 SDVARAMOILSGKRLGLAIVDDGTGYSSSLYKLPFIDVLDKIDRSFVGLPRGEDDA 1347  
 OY 801 AIIQVVIDLANGLLETVAEIGSEAOLORLQKMGCHGGGFTLRLPAEAMMTYLYYP 860  
 DB 1348 QIARIIIMASHLNLVIAEGVESQALDFLREHGCDEVQYILFRPMAP- QFGMLYAS 1406  
 OY 861 QIL 863  
 DB 1407 DVL 1409

RESULT 2  
 ID 09A3V7 PRELIMINARY; PRT; 748 AA.  
 AC 09A3V7;  
 DT 01-JUN-2001 (TREMBLREL. 17, Created)  
 DT 01-JUN-2001 (TREMBLREL. 17, Last sequence update)  
 DT 01-DEC-2001 (TREMBLREL. 19, Last annotation update)  
 DE SENSORY BOX/GDEF FAMILY PROTEIN.  
 GN CC3094.  
 OS Caulobacter crescentus.  
 Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;

OC Caulobacter.  
 OX NCBI\_TaxID=69394;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=21173698; PubMed=11259647;  
 RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
 RA Pockock I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
 RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
 RA Kolonay J.F., Smit J., Caven M.B., Khouri H., Shetty J., Berry K.,  
 RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,  
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.,  
 RT "Complete genome sequence of *Caulobacter crescentus*."  
 RI Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 DR EMBL: AE005972; AAK25056.1; .  
 DR TIGR: CC3094; .  
 DR InterPro: IPR001633; DUF2.  
 DR InterPro: IPR000160; DUF9.  
 DR InterPro: IPR000014; PAS.  
 DR InterPro: IPR000700; PAS-assoc\_C.  
 DR Pfam: PF00563; EAL; 1.  
 DR Pfam: PF00990; GDEF; 1.  
 DR SMART: SM00267; DUF1; 1.  
 DR SMART: SM0052; DUF2; 1.  
 DR SMART: SM00091; PAS; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 748 AA; 81907 MW; 5809F32970834CE9 CRC64;

Query Match 13.3%; Score 882; DB 16; Length 748;  
 Best Local Similarity 36.7%; Pred. No. 9.5e-53;  
 Matches 203; Conservative 100; Mismatches 190; Indels 60; Gaps 9;

OY 354 GSTLLREKESLYKWMACKRGIDRNNIIPRLSEAMEETOKLVPTWNRSEKRLAQVASTQ 413  
 DB 226 GGMALTRTKTALLNAEG-----RAEYLLGVSEDAER-----KRAEQIARLANH-- 280  
 OY 414 LYMATQOEVTRILITQOTAYDPLTQLPNMTIFNRQLTALLDALYEGKAVGVLYVIMDRF 473  
 DB 281 -----YDPLTLPNRYLFQKSLGALARRSKGDALAHVFYDLDRF 321  
 OY 474 KRINESFGHKTGGGLLOEVADELNOKLSPLAAYSLLSRMHGDGFTILLTQISDNQEMIP 533  
 DB 322 KTVNDTLGHPGLDALLKTAERLR---GCVREGDTVARLGDEFAIVQTGLDSDNGATR 377  
 OY 534 LCERLSTFOEPFFQOGPIYLTASMGISTAPYDETAESLKFELIALTRAKCOGKNTY 593  
 DB 378 LAARIVEAMARFELQGHVYIGASVGSVLAPTDGDDDELKRDMLYRAKADGRAY 437  
 OY 594 QFYRPODSAPMLDRLTLESDLRQALTNQEFVLYEQPYVALDTGKLLGVEALVRWQHPRLG 653  
 DB 438 HFERAMDEQLOARALRLTLESDLRRAALQAGEFELEFQPLVHLGDERVTGCEALLRMHPRBG 497  
 OY 654 QVAPRVFPLAEELGINLNGWVLETACATHQHFRETGRRLRAVNVISARQODEKVL 713  
 DB 498 MVSFADFIPLAEELGLIYQGEWVLRACADAAMN---PEHVRLAVNLSPAQFDRGLV 553  
 OY 714 NSVLECLKRTGAPPEDELEITESLIMEDIGTVVLLHRLREGEYQVAIDFGTGYSSLS 773  
 DB 554 RTVSALAAASGLPRAORELEITESVLIQDSQANMIMLHDLKALGVRISIMDEFGTGYSSLS 613  
 OY 774 ILKQLPIHRLKIDKSFVNDLNEGADTAIIQYVIDLANGLLETVAEIGSEAOLORLQK 833  
 DB 614 YLRSPFPKIKIDQFVNDLHDSQAMAIKAVLDLGSAGMVGVTAEVETQAOALDALRQ 673  
 OY 834 MGCHIGGQYFLTRPLPAAMMTYLY-----PQIID-FGPIPLPKVALPETET--- 881  
 DB 674 QGCAEIOGYFTISRPAPASEIAKMLGVEGRADLGAPSVLSPIGANPPPO-AGOEVRTAPS 732  
 OY 882 -----EAGQNGV 889  
 DB 733 SPVARTGEGSRG 745

RESULT 3  
ID 055576 PRELIMINARY; PRT: 1244 AA.  
AC 055576;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 141.7 KDA PROTEIN.  
GN SLR0359.  
OS *Synechocystis* sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; *Synechocystis*.  
OX NCBI\_TaxID=1148;  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN-PCC 6803;  
RA Tabata S.;  
RL Submitted (Aug-1995) to the EMBL/GenBank/DBJ databases.  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE-96127529; PubMed-8590279;  
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,  
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,  
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
RA Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
RT *Synechocystis* sp. strain PCC6803. II. Sequence features in the 1 Mb  
RT region from map positions 64% to 92% of the genome.";  
RL DNA Res. 2:153-166(1995).  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE-97061201; PubMed-8905231;  
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
RA Miyajima N., Hirosewa M., Sugita M., Sasamoto S., Kimura T.,  
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,  
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
RA Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
RT *Synechocystis* sp. strain PCC6803. I. Sequence determination of the  
RT entire genome and assignment of potential protein-coding regions.";  
RL DNA Res. 3:109-136(1996).  
DR EMBL: D63999; BAA10080.1;  
DR InterPro: IPR001633; DUF2.  
DR InterPro: IPR001603; DUF9.  
DR InterPro: IPR003018; GAF.  
DR InterPro: IPR001610; PAC.  
DR InterPro: IPR000014; PAS.  
DR InterPro: IPR000700; PAS-assoc\_C.  
DR Pfam: PF00563; EAL; 1.  
DR Pfam: PF01590; GAF; 1.  
DR Pfam: PF00990; GDEF; 1.  
DR Pfam: PF00785; PAC; 4.  
DR Pfam: PF00989; PAS; 3.  
DR SMART: SM00267; DUF1; 1.  
DR SMART: SM00052; DUF2; 1.  
DR SMART: SM00065; GAF; 1.  
DR SMART: SM00086; PAC; 4.  
DR SMART: SM00091; PAS; 4.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 1244 AA; 141661 MW; 23D80BD015ECF559 CRC64;

Query Match 13.1%; Score 865; DB 16; Length 1244;  
Best Local Similarity 31.0%; Pred. No. 3, 3e-51;  
Matches 259; Conservative 141; Mismatches 286; Indels 150; Gaps 28;

QY 145 VPMMDQ-----QMGIMAV-----HHSKPRRF-----TEQETMAL----- 177  
DB 475 IPVADQAEITLTPDGSMMVAEMKVSPIHMQGKAFVLSRDITRHOARLALAESEK 534  
QY 178 -----LSKEVSLATTSQSLSRQVHQ-----OVQEALVORLTETVAQVQYDREFTWQ 223  
DB 535 KYRIYELTSEGIWLDQDQQTTFANQGLADMLGYSVQELLENKNTAFVLVTHHPESON 594

QY 224 YAEIV-----GQAVEA-----DGAVLYIADLTGSAQHTQWNLRFDMKMW-- 265  
DB 595 SHQKTLQSPFCVLPNHCQVYDVQFRRDGSVLM-----GLVRSAMYD---QMGVYRG 645  
QY 266 ---LETSLMGELMRGQSAAMEP-----MAAYOS-TWEKRPFTSVAPLPTTCVPHGYT 316  
DB 646 ELAMLDITRKSAEQALSSSEORLEIGLSIDQVWASADAVSPALYLPPTTAMVYGOS 705  
QY 317 LGELEQRSD-W---IAPRESLSAENFQSFLLVPLAADQCVWGSLLILRKESLVKHWAGK 372  
DB 706 LEVQSQNWFQVHNGRDLLEHYQLL---MENQ-----TELYRIVQPGSGKE 754  
QY 373 RGIDRNIILP-----LSEWAMEETOKLVTNNRSERKLAQVASTOLYATITQCVTRL 426  
DB 755 RWLFRRSQVLVDEHQPLRIDSIDS-----DITERKLA-----AEK 790  
QY 427 ITQQTAVDPLOTLPNMTTFEKRQTLALDLALYEGKM-VGLVYIAMDREKRNESFGHKG 485  
DB 791 LHYNANHSDLTNLPNNSMFLDRGLHALQRLRRDLRFVAVFLDLDFGFKLINDSLHSGC 850  
QY 486 DGLQEVADRLNQLSPLAAYSPILSRWHDGFTILLTQISDNOEMIPLCERLSTFOEP 545  
DB 851 DLLQGIARHLRQCLRP-----EDTLARLGDEFTMLEIENTCEDVYAVAQRIHQELQRP 906  
QY 546 FPLQGPITYLTASMGIS-TAPYDGETAESLTKFAEIALTPKACQGNFTQYFRPOSAPM 604  
DB 907 FNLNGQEIFNTSIGIALNHPHYGH-PQDVLRADPTAMYKAGKRAYAFNQTMHNA 965  
QY 605 LDRLTLESRLQALTNQEFVLFQPOVALDPTGKLIGVLEALVYRMOHNPRLGQVADVFPLA 664  
DB 966 VQRLQRENDLRAIDRELQHLQYVCLKTGQLOGEALVYRMOHNPREGILDEEFYALA 1025  
QY 665 EELGLINHLQWVLETA---CATQHFRETERGRRLMAVNISRQFODERKWLINSVLECL 720  
DB 1026 EETGLVPMGDWILMEASRQILDELKQSFQLS--HLQVSINVSRRQRLDKTVDTEL 1083  
QY 721 KRQMPPEDELEETESLIMEDIKGYVLLHRLREBSQVYAIIDFGYGSLSLTKQLPT 780  
DB 1084 SSTNLAPQDLKLEITESTSLDNLNLADVLKSRQRNIQISLDFGYSLSYLHNEPI 1143  
QY 781 HRLKIDKSFYNDLNEGADPAIIQYVIDLANGNLEFVAGISEAQLOKMGCHLQG 840  
DB 1144 NTIKVDKSPVNTMEPNNONTAIYHTVLAHTGLDVIAGIETERTLTLQMLHLCGDAQ 1203  
QY 841 GYFLTRPLPAEAMTYLVYQQLIDFGTPPLPVVALPETETEGAGQGNVGDRLPNS 896  
DB 1204 GYFARPIPSIEDLDFL-----TIQIPMSVPAV-TECH-----GPRKEPFS 1244

RESULT 4  
ID 0915W1 PRELIMINARY; PRT: 1245 AA.  
AC 0915W1;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL PROTEIN PA0575.  
GN PA0575.  
OS *Pseudomonas aeruginosa*.  
OC Bacteria; Proteobacteria; gamma subdivision; *Pseudomonadaceae*.  
OC *Pseudomonas*.  
OX NCBI\_TaxID=287;  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 15692 / PA01;  
RX MEDLINE-20437337; PubMed-10984043;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,  
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber L.L., Goltzer L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Polger K.R., Kas A., Larij K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reiter J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an





```

OY 577 FAEIILTAKKOGKNTQOFEYBPQDASMBRLTEJSLPQALNFOEVLKFOVALDGT 636
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 487 NADAMATRAKELGRNNNQVLFTEEMAAARHKELOQSEELREALAREPFLIFQOMSLTG 546
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 637 KILGVEALVEMOHRPLGOVADVEFIPLAEILGINLIGOWVLETACATHOHPFRETGRIL 636
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 547 RIFAFAELLRMOHVRGRIISPAFIRPLAEFGLIVPGDWLRLTAC-----ROL 595
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 697 R-----MAVNISARFOFOEKMLNVSLBCTKRTGMPREDLELITESLIMEDIK 745
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 596 KAKODAGILPLIYSVNSAROFERRNNARAAVAVLEETCLEARTIELBELTESLIMQDPG 655
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 746 TVVLLHRLREGVQVAIDDFGTGSSLSILKQPIHARKIDTSFVNDLNEGADTAIIQY 805
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 656 AIAATMHAEIGAIVHAILDDFCTGSSSALTKRFVVRRIKIDRSFVTDIPHADDMATTSA 715
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 806 VIIDLANGLNETVAEGIESEAOIOLKMGCHLGQGEFLRPLPAE 851
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 716 IISIAOKIGLKVIAEGVETAEQVLELCKSGCGTIGTGFFSOPLEGE 761
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT	6			
0916K5				
ID	0916K5	PRELIMINARY;	PTT;	760 AA.
AC	0916K5			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	HYPOTHETICAL PROTEIN PA0285.			
GN	PA0285.			
OS	Pseudomonas aeruginosa.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;			
CC	Pseudomonas.			
OX	NCBI_Taxid=287;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 15692 / PA01;			
RX	MEDLINE=20437337; PubMed=10984043;			
RA	Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,			
RA	Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,			
RA	Gader R.L., Goltzy R., Tolentino E., Westbrook-Wadman S., Yuan Y.,			
RA	Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,			
RA	Reizer K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,			
SA	Selzer J., Sailer M.H., Hancock R.E.W., Loy S., Olson M.V.; an			
RT	"Complete genome sequence of Pseudomonas aeruginosa PA01, an			
RT	opportunistic pathogen.";			
RL	Nature 406:959-964(2000).			
DR	EMBL; AEO04466; AAG03674.1; "			
DR	InterPro: IPR001633; DUF2.			
DR	InterPro: IPR000160; DUF9.			
DR	InterPro: IPR001610; PAC.			
DR	InterPro: IPR000014; PAS.			
DR	InterPro: IPR000700; PAS-assoc.C.			
DR	Pfam; PF00990; GDEF; 1.			
DR	Pfam; PF00785; PAC; 2.			
DR	Pfam; PF00989; PAS; 1.			
DR	SMART; SM00267; DUF1; 1.			
DR	SMART; SM00052; DUF2; 1.			
DR	SMART; SM00086; PAC; 2.			
DR	SMART; SM00091; PAS; 2.			
KW	Hypothetical protein; Complete proteome.			
SO	SEQUENCE 760 AA; 86380 MW; 049AFDID6E222E05 CRC64;			
Query Match	12.66;	Score 830.5;	DB 16;	Length 760;
Best Local Similarity	33.36;	Pred. No. 3.8e-49;		
Matches 238;	Conservative 104;	Mismatches 248;	Indels 125;	Gaps
OY	214 QYDGPETWQVAL--ETVGAIVEDGAVLYTAPDLTGSVAOHYOWNLRFPMGN--TWLETS 269			
Db	120 ELGDPTFTWKKRLNPEEATVYLEARNHL---GSLTNDLHIYR--LRHKGDDVRLWHS- 173			
OY	270 LWOELMRGQPSAAMEPMAAVQSTWEKRPPTSVAPLDPPTNCVPHGYLGELEQ----- 322			

Db	174	-----ROR-----	VLNDALGKPLHYGVA-----	RDITLQRLKEDHLQAA	209
Qy	323	-----	-----RSDMT-----	APPSLSAENFQSR--	LIV 344
Db	210	AVFDSTREGVLTDAQAVIVHVNPSFERITGVRSEDVGLKTPALIRSGRODAQAYFORLWL	-----	-----	269
Qy	345	PLADQQQWVGSLLILRKRSLYVKHMAGRGIDRNNILPRLSFEAMEETQKLVPMWNSER	-----	-----	404
Db	270	ALREDQVMSGELWMNRKRSGETIYPOWLHRAV--RNDGQQLTHYGVSDD--LSSIKRSEN	32525	-----	
Qy	405	KLAQVASTQLYMAITQOQFVTRLITQQAADPLTQAPMWIIFNRQLTALLDALYEGKMVG	4646	-----	
Db	326	ELDFLAAH-----	-----HDSLTGLPVRVILLRETRIEQALENGKDRTVAGA	3646	
Qy	465	VLYVIAIMPKFRINESFGHKTDGDLGQAYADRLNKLSPLAAYSPLLSRMNDGFTLLTQ	52424	-----	
Db	365	LLLLIDHFKHINSLGHTTDMLLKEVSKRLQHOLBERC--LISRLGGDEFALIV-E	419	-----	
Qy	525	ISDNOEMIFLCEERLSTQOEFPFLQOGQPIYLTAFASMGISTAPYDETAESLKFPAELATR	584	-----	
Db	420	NDDPEAVARLSQRILLDGSNAFEDHICPIYISASLGVSIVPEASDVDDHLMQNDALAPQ	479	-----	
Qy	585	AKCGCKNTYQFYRPDQSAFMLDRLTLESDDLQALTFNQEYVLVFPQYVALDTGKLLYEAL	6444	-----	
Db	480	AKDSGRNAYAEYTRLLTARARAHQVESALRHALEHDELRYHYQPVHDLASGRIVGESL	539	-----	
Qy	645	VRMOWPRLGOVAPDVFPLPLAEELGLINHLGOWVLETPCATQHOFRENGRLR-MAVNIS	703	-----	
Db	540	VRMOWPREGGLVPPGGEFVYVAEEGCLIALDNWVLYKRCROMRE-WOORGYELEFVAIVNS	598	-----	
Qy	704	ARQFQDEKWLNSVLECKRTGMPPEDELETESLAMEDIKIGVALLHRLREEGVOYAI	763	-----	
Db	599	SRLEFRGELERRIANALLESGLPRLYTELEVETESAVMEDEFOSSINTLCRLILGVNAID	658	-----	
Qy	764	DEGCGYSLSLTKQPLPHRLTKIDMSFVNDLNEGADPAITQIYVDIANGLNTLEFVAEGIE	828	-----	
Db	659	DFGYTSLSLMRKRLRPLPVKLLIDGFFVAGLGAVDADAIAIAVALQSMGLRVAAIGIE	718	-----	
Qy	824	SEAOQLQRQKMGCHLGQGYFLTRPLPEAMMTYLYYPOILDFGTPPLPKVALPE	878	-----	
Db	719	HQDQALFLREHGDCFGQGWYGRQPANALNF-----	DLPPALPE 759	-----	
RESULT	7				
Q9HYT3		PRELIMINARY;	PRT;	783 AA.	
AC	Q9HYT3:				
DT	01-MAR-2001 (Tremblrel. 16, Created)				
DT	01-MAR-2001 (Tremblrel. 16, Last sequence update)				
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)				
DE	HYPOTHEITICAL PROTEIN PA3311.				
GN	PA3311.				
DS	Pseudomonas aeruginosa.				
CC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;				
OC	Pseudomonas				
OX	NCBI_TaxID=287;				
RN	SEQUENCE FROM N.A.				
RP	STRAIN-ATCC 15692 / PA01;				
RC	MEDLINE=20437337; Pubmed=10984043;				
RA	Stover C.K., Pham X.-O.T., Ervin A.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,				
RA	Hickey M.J., Brinkman F.S.L., Hutnagle M., Westbrock-Madhan S., Yuan Y.,				
RA	Gardner R.L., Goltzy L., Tolentino E., Westbrock-Madhan S., Yuan Y.,				
RA	Brody L.A., Coulter S.N., Folger K.R., Kas A., Larbig I.T.,				
RA	Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,				
RT	"Complete genome sequence of Pseudomonas aeruginosa PA01, an				
RL	opportunistic pathogen."				
RL	Nature 406:959-964(2000).				
DR	EMBL; AE004753; AACG06699.1; -				
DR	InterPro; IPR001633; DUF2.				
DR	InterPro; IPR000160; DUF9.				

QY	565	AKGCSKNFYQYRRQDSAPMLDRITLTSLEDRQALATNQEENFYVLPPOQVAILDRGKLLGEAL	644
Db	440	NDDEEVAARLISQRIILIDGNAFEDHICOPITYISASLIGVSLPEPDSVDHDMOHADALFQ	479
QY	565	AKGCSKNFYQYRRQDSAPMLDRITLTSLEDRQALATNQEENFYVLPPOQVAILDRGKLLGEAL	644
Db	480	AKDSGRNNAYATYTRYLTKARAAHVQESALHHALEHDLRHVYQPVHDLASGRIVGEVL	538
QY	645	VRMHQPRLAGOVAVDVEFIPLAEEELGILNMGVLETFACATQHOFRETRGRLR-MAVNIS	703
Db	540	VRMHQPRGGLVPPGFEFVPAEECGILALDNWVLKRCRQMR- WQGRGVELEFVAANVS	598
QY	704	ARQPDDEKWLNSVLECLKRTGMPEDDELETETESLAMEDIKGTVLLHRLRESEVOAID	766
Db	599	SRLEFRKGLEERIANALAEESGLEPYLELEVEYESAVMEDEPOSILNLCLRLILGVLNAID	658
QY	764	DFGTGYSLSLTKQPHIRLTKLSEFVNDLINESADPAIRAIQYVLDLNGLLETFVAGIE	822
Db	659	DFGTGYSLSLMLKRLPIYKILKIIDQGFVAGLPGCAVDDALAIAYALQSMGLRVVAGIE	718
QY	824	SEADQLRQKMGCHLGQGYFLTRPLPAEAMTYLYRQIIDFGFTPLPVALPE	878
Db	719	HODQALFREHGCDPFGQGYWGRQPAEALRF-----DLPPVALPE	759
RESULT	7		
Q9HYT3		PRELIMINARY;	PRT; 783 AA.
AC	Q9HYT3;		
DT	01-MAR-2001 (TREMblrel, 16, Created)		
DT	01-MAR-2001 (TREMblrel, 16, Last sequence update)		
DT	01-DEC-2001 (TREMblrel, 19, Last annotation update)		
DE	HYPOTHETICAL PROTEIN PA3311.		
GN	PA3311.		
OS	Pseudomonas aeruginosa.		
CC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;		
OC	Pseudomonas.		
OX	NCBI_TaxID=287;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-ATCC 15692 / PA01;		
RX	MEDLINE=20437373; Pubmed=10984043;		
RA	Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,		
RA	Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,		
RA	Gardner R.L., Goltzy L., Tolentino E., Westbrook-Ladman S., Yuan Y.,		
RA	Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,		
RA	Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,		
RT	*Complete genome sequence of Pseudomonas aeruginosa PA01, an		
RL	opportunistic pathogen..;		
RL	Nature 406:959-964(2000).		
DR	EMBL; AE004753; AAC06699.1; -		
DR	InterPro; IPR001633; DUF2.		
DR	InterPro; IPR000160; DUF9.		

DR InterPro: IPR000215; Serpin.  
DR Pfam: PF00990; GDEF; 1.  
DR SMART: SM00267; DUF1; 1.  
DR SMART: SM00052; DUF2; 1.  
DR PROSITE: PS00284; SERPIN; UNKNOWN.1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 783 AA; 86815 MW; 7D81B379356430F CRC64;

Query Match 12.3%; Score 811.5; DB 16; Length 783;  
Best Local Similarity 29.7%; Pred. No. 8.5e-48;  
Matches 239; Conservative 119; Mismatches 285; Indels 161; Gaps 19;

QY 160 HSKRRP-----TQEWETALSKSE-----VSLATQSLRQVHOQ 198  
DB 43 HRRRRVRFVQGDFTAMDQGLRFLGESPYGVYQNCYSPSLVALAFVLAAGTALD 102  
QY 199 QVQVALVORLETTVAQYGRPEWTQ-----VALEIV 229  
DB 103 -----MVERVNSLS-----HPLMOWIGAFCLGSGIMATHEVAMLAFAHPIALRYDLPIT 153  
QY 230 GOAVBADGAVLYIAPDLT-----GSVAOHYOMNLFDMGNWLETSL 270  
DB 154 GLSLIAVAASYLTWMTARPRFGLPCLLACCGIGIAAHY----- 197  
QY 271 WOELMRGQPSAAMEPMAVQSTWEK-----RPTSVAPLPPTKCVPHGYTLG 318  
DB 198 -----TGMAMRSVATQYQPSLFAISVLIAGNAFTALAVP-----YLRG 239  
QY 319 ELEORSMDIAPRESL-----SAENFQSP-----LIVPLAADQWGSILILKREKSLVKH 368  
DB 240 RRSARYRWKLIALSLILGALIAAHFTGMALVLSVPACT-----PLEQASADSLRIG 293  
QY 369 WAGKRGIDRRNLPRLSEAMEETQKLVPTWNRSEKRLAQVASTOLYMAITQ-QEVTRLI 427  
DB 294 WL--TGVLASATAACGIAAMSEKOR-----ERRLSE--NSRYNALNLQDHAHSL 341  
QY 428 TQQTIVYDPLTQPMIIFNRQTLALLDALYBGMVGLVIMDFKRIENEFGRKTGDG 487  
DB 342 ROMARYDLSLTGLQNTAANEVFNQHLNCRKGLKGLAVWFLLDHFKRINSLGHDGQ 401  
QY 488 LLOEVAIDLNOKLSPLAAYSPILSRWHDGFTILITQISDNQEMPLCERLSTFOEPFE 547  
DB 402 LKTIYSERIRSVLRD-----SDVVARFADGECVLAIDLQDHAH--LSQRLMKKKEPIA 456  
QY 548 LQGDPIYLTASMGISTAPYDGETAELSLKFAEIALFRACCKGKNTYOFYRPQDSAPMLDR 607  
DB 457 LDGRTLVMTASGVSLYPNDGECCELLKNGALHQSACGRNNAOFFSRQLIVRATQE 516  
QY 608 LTLBSDLROALTNQEFVLYFOQVADLPDCKLLGVNRMQHPRLRGQVAPVFTLABEL 667  
DB 517 LQMBELRQALRDQDLHYQPTIALADGEVHQLEALVRMRPPTGGLGPRDFIGLAEAN 576  
QY 668 GLINILGQVLETAACATQHFFRETRGRRLMAVNISARQODEKMLSVLELCKRTGMP 727  
DB 577 GMIOQLDWMVLRACRDLRSLHAGHERLRYAVNCCASNLRASLVDEVRNHLDEAGLAA 636  
QY 728 EDLELEITESIMEDIKTVVLLHRLREGVQVADIDEGTGYSSLSLKQDPIHRLKIDK 787  
DB 637 CFLELEVETEDALNIDQITIPLERLRELGVSLSIDFGTGYSSLAYLRRLDLALKVDR 696  
QY 788 SFVNDLNEGADTALIQVIDLANGLNEYAEGIESERQOLRQKMGCHLGQGYFLTRP 847  
DB 697 SFINDIPASQDMEIAQAIIMAAQCLHLKVVAAEGVETPOALAFIRENHCELVGYSRSP 756  
QY 848 LPAEAMTYLYYPOIDGCPTRPL 871  
DB 757 LPLALEEFL--RAYRPDAAPL 777

RESULT 8  
Q91310 PRELIMINARY: PRT: 685 AA.

AC Q91310;  
DT 01-MAR-2001 (TRENBLREL. 16; Created)  
DT 01-MAR-2001 (TRENBLREL. 16; Last sequence update)  
DT 01-DEC-2001 (TRENBLREL. 19; Last annotation update)  
DE HYPOTHETICAL PROTEIN PA1727.  
CN PA1727.

OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Gaeher R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lathig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Salier M.H., Hancock R.E.W., Lozy S., Olson M.V.,  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
RT opportunistic pathogen.";  
RL Nature 406:959-964(2000).  
EMBL: AE004599; AAC05116.1;  
DR InterPro: IPR001633; DUF2.  
DR InterPro: IPR000160; DUF9.  
DR Pfam: PF00990; GDEF; 1.  
DR SMART: SM00267; DUF1; 1.  
DR SMART: SM00052; DUF2; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 685 AA; 74378 MW; 4BF74F5CF49A40C7 CRC64;

Query Match 12.3%; Score 811; DB 16; Length 685;  
Best Local Similarity 34.4%; Pred. No. 7.4e-48;  
Matches 196; Conservative 107; Mismatches 179; Indels 88; Gaps 13;

QY 340 SFLIVPLAADQ-WV-----GS--LILKREKSLV-----KHWAK----- 372  
DB 145 SILIVIVASGALMTAFRLRHSRRLVLRGALVMGCAIVGMHYTMAAOPPLGSCF 204  
QY 373 -----RGIDRRNLPRLSEAMEETQKLVPTW-----NRSERKLAQVASTOLY 415  
DB 205 GAAGGCID-----NGMLAVLYITTLAVIAIALYVSLSRLER--TSVATSLA 253  
QY 416 MAITQVYRLTQGTAVDPLTQPLPMIIFNRQTLALLDALYBGMVGLVIANDRKR 475  
DB 254 RA-----NRELQALHDNLTKLPNRLDRLQALIQQAIRDDRRAVLFMDLGRKA 307  
QY 476 INESFGHTGQGLQEVADRLNOKLSPLAAYSPILSRWHDGFTILITQISDNQEMPLC 535  
DB 308 VNDAGHNLGDLLEVAERIRANVRA---QDTIARLGDEF-VLLIAREPADAAITLA 362  
QY 536 ERLSTFOEPPEFLQGPITYLTASMGISTAPYDGETAELSLKFAEIALFRACCKGKNTYOF 595  
DB 363 EKLAKRISQPIQIISHEVRISASIGIALYPRDQGTREHLMINADAMNAADQGRNGYCF 422  
QY 596 YRPQDSAPMLRRLTLBSDLROALTNQEFVLYFOQVADLPDCKLLGVNRMQHPRLRGQV 655  
DB 423 FESSMNAQAQQLDHLRLQALERQVLYHYQPKVLAHPNPMGYEALLMEHPHGLI 482  
QY 656 APDVFIPLAEELGLNHLGQVLETAACATQHFFRETRGRRLM-----AVNISA 704  
DB 483 TPQGFPLAEKTYGLIVQGEVWVDEAC-----RQRLMIDGGHADMIANVLSA 531  
QY 705 RQODEKVLNVLLECIRTGMPREDLEITESIMEDIKTVVLLHRLREGVQVADID 764  
DB 532 LQFAHAGLVDSVRNALLNHSLEPSHLILEVRESAMRADASIVILRQSLAMGVGISID 591  
QY 765 FGTGYSSLSLKQDPIHRLKIDKSFVNDLNEGADTALIQVIDLANGLNEYAEGIES 824  
DB 592 FGTGYSSLSLYKRLPASLTKIDRGFINELANDSDDAIYSAIYALGRTLNKTYAEGVET 651

```

QY 825 EAQIOLKMGCHLGGYFLTRPLPAEAMM 854
DB 652 EAQOEFLTRIGCNLSLGGFLGRPMPEOILL 861

RESULT 9
Q9HUF2 PRELIMINARY: PRT: 899 AA.
AC Q9HUF2;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHEITICAL PROTEIN PA5017.
GN PA5017.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Hickey M.J., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Martener P.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Laidig K., Lam R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RT Nature 406:959-964(2000).
RL EMBL, AE004914; AAC08402.1; -
DR InterPro: IPR001633; DUF9.
DR InterPro: IPR000160; DUF9.
DR InterPro: IPR003018; GAF.
DR InterPro: IPR001610; PAC.
DR InterPro: IPR000014; PAS.
DR InterPro: IPR000700; PAS-assoc.C.
DR Pfam: PF01590; GAF; 1.
DR Pfam: PF00990; GDEF; 1.
DR Pfam: PF00785; PAC; 1.
DR SMART: SM00267; DUF1; 1.
DR SMART: SM00052; DUF2; 1.
DR SMART: SM00065; GAF; 1.
DR SMART: SM00086; PAC; 1.
DR SMART: SM00091; PAS; 1.
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 899 AA; 101321 MW; 925CIFAG36731602 CRC64;

Query Match 12.3%; Score 810.5; DB 16; Length 899;
Best Local Similarity 27.5%; Pred. No. 1.3e-47;
Matches 248; Conservative 151; Mismatches 289; Indels 215; Gaps 27;

QY 5 RSLDEFLRNVLNKK-----FHALT---LRETLQVYIEEARIFLGVDRVVIYKRA 50
DB 160 RSODHLDHLKRSRTQNLIVRLANRHYLSDDLLEAQLTQACEAVTARACIWRLL 219
QY 51 SDGSGEVL-----EAVNRALPSILGLHFVEDIPQAREELGNQRKMIADV 98
DB 220 DDQRLAENVYVRRLDDQYKEKPSIDASRYPAL-----EAVNSGALIDAH 264
QY 99 VAHRKKKSHHSIGRISPTSHSNGHYTVDSCHIQYILANGVLSLTPVMQDOGLMIVA 158
DB 265 NAORPRTIOEL-----YKDYLRPLGVNALIDATIRIGEVVGYLC 304
QY 159 VHSRPRRTQEWETMALTSKEVS-----LATOSQLSQVNOQVOQVLEAVLRLETTAAO 214
DB 305 LEHAENEM-----WQ-----SDEIAPAGELADQYAOVLMNHRVSSSL----- 345
QY 215 YGDRRETFQVALVYVGOA-----VEADGAVLYIAPDLTGSVAQYQNNLRPDMGNMLETSL- 270
DB 346 -----HLQRAVVEGASASFLILDRDGVVEYVNPST-SITQVSADEVR-----NRRLSELP 395

```

```

QY 271 ----WQELRGQPSAAMEPMAAVOSTWEKRPBFTSVAPLPNTNCPHGTYLGELEQRSDM 326
DB 396 ALENISELLFIDARSALTQ-----QNSMQ-----GEFRSR----- 424
QY 327 IAPPELSAENQSFILVPLADQOVGSIILIRKEKSL--VKHMGKKGIDRRNRLPRL 384
DB 425 -----RKNHEPY-----W-GQLSLSKVDDIGELTHYIG----- 452
QY 385 SPEAMEEQKLVPTWNRSEKLAQVASTQLYMAITQOFTRLITOOTAY-DPLTOLPMNI 443
DB 453 -----IYEDITONKLAQOHTKLAIDNLTGIA---- 480
QY 444 INRQLTLLALDLYBG---KNWGVLYIAMDRKRNESFGHKTGGDGLQEVADRILNQL 500
DB 481 --NRHYFIALERLESSDRPLSLILVINDNFKRINDSLGHTQDGKLVSLARRLSCL 538
QY 501 SPLAATSPILSRWHDGFTILL--TQISDNQEMIPICERLSTFOEPFLQOPITYLTAAS 558
DB 539 GGGAT---LARPASNEFVLLDDTAVERGES---IAQVLMHLDKPLFVDNQLINITGS 591
QY 559 MGISTAPYDGETAESILKFAETALTRAKCGKNVYGFYRPOSAPMLDRLTLESRLRAL 618
DB 592 IGLASAPQHCQDPQTLMTAGLALHKAKANGHOVYFTEALTAESTYLFPESNLRRL 651
QY 619 TNOEFLYRPOVALDTGRLGVEALVRWQHPRLGQAVDVTPLAEELGLNHLGQWVL 678
DB 652 AQNELAVHQPRLCLRSGLGLEALLRMQHPEKMIKRPDRFISVAEFTGLVPICKWVI 711
QY 679 ETACATHQFFRETGRRLMAVNIASROFODEKVLNSVLECLKRTGMPDELETTESL 738
DB 712 REACRQARELABRAGLGELOIAINLSPKPTDPLDVESIAIILHEENIPASOLELETESL 771
QY 739 NMEDIKGVVLLHRLREEGVOVAIDPFGTGSLSLTKOLPTRLKIDSFVNDLNEGA 798
DB 772 LLDADDTTQOELRLKSLGLTLTAMDFTGYSLSLTKLFPIDYIKIDISFIKIDISD 831
QY 799 DTAITQYIDLANGLNLEFVAIGISEAQOLRLQKMGCHLGGYFLTRPLPAEAMT-YL 857
DB 832 DMEITSAVIYAMHNKLKVAAGVESAEOLAFIRNRCDIGOGYLDPRDIPSLNTLSL 891
QY 858 YYP 860
DB 892 RYP 894

RESULT 10
Q9HW35 PRELIMINARY: PRT: 687 AA.
AC Q9HW35;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE HYPOTHEITICAL PROTEIN PA4367.
GN PA4367.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Hickey M.J., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Martener P.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Laidig K., Lam R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RT Nature 406:959-964(2000).
RL EMBL, AE004852; AAC07735.1; -

```

DR InterPro: IPR001633; DUF2.  
 DR InterPro: IPR000160; DUF9.  
 DR Pfam: PF00563; EAL; 1.  
 DR Pfam: PF00990; GDEF; 1.  
 DR SMART: SM00267; DUF1; 1.  
 DR SMART: SM00052; DUF2; 1.  
 DR Hypothetical protein; Complete proteome.  
 KW SEQUENCE 687 AA; 77564 MW; 4907154941BC53DE CRC64;  
 SQ

Query Match 12.2%; Score 809; DB 16; Length 687;  
 Best Local Similarity 35.8%; Pred. No. 1e-47;  
 Matches 226; Conservative 85; Mismatches 246; Indels 72; Gaps 16;

QY 266 LETSIQMOELMR-----GQPSAMEPMAAVOSTWEKRPPTSVAPLRPTNCV----- 311  
 DB 75 LEGLFQHEAVRQASIGHNP-----EPMLA-----EKSRLPDL-----PTRWLTDPILGQERT 122  
 QY 312 -----PHGYTIGLEQRSDMTAPRESLAENP--QSFLIVPLAAQOQVGSILILR 360  
 DB 123 PSIRLIGRPPEYSEYGDLTITLD-TAP-----YGENVTTSITLITISGLIRALAMGLVLF- 176  
 QY 361 KEKSLVKNHAGKRGIDRRNLRPL-SFEAMEETOKLVPT--WNSEKRLAQVASTOLYM 416  
 DB 177 ----LVYHNMMLTKPLSK--IIEHLVSINDRPSQHQLPLKGNHERNELGMLVTTANQLLA 230  
 QY 417 AITQGFVTRL-----ITQGTAYDPLTQLEPNIENRQLTALLDALYES-----KNVGLV 466  
 DB 231 STESNHLRREAEDMLRTISQYDFLTGLP-----NRQLLOQOOLDQILDDAGRQRRVAVL 285  
 QY 467 VIANRFRKRNESFGHKKGDDILOEVADRNLNOKLSPLAAYSPLLSRMHGDDFTILLTQIS 526  
 DB 286 CIGLDFEKGINQYVQLGDOLLIALADRLRGHSARLGS---LARLGGDQFALVQADIE 341  
 QY 527 DNOEMIPCELRSLTFQEPFLOGQPIYLVTASMGISTAPYDGETAESLFAEIALTRAK 586  
 DB 342 QPEYAEALQSLDLDLENAPFEIDQNEHVRKATIGITLPEDESETTEKLQKAEQMTLAK 401  
 QY 587 CQCKNTQYFYPQDSAPMLDRLLTESDLRQALTNQOEVLVYFQVAYALDTGKLLGEALVR 646  
 DB 402 TRSRNRQYFYIYASVDSKMRRELEKDLRDLRGHEHLNVYQVYDRHVGVGEALLR 461  
 QY 647 WOHPRLGQVARDVFIPLAEELGLNHLGQWVLETACATHQNHFFRETGRRLRAVNI SARQ 706  
 DB 462 WOHPRHGVPRLDFTPLAEQNGSIFSGEWMVDQACROLREHNDGFDDLRAVNLSTVQ 521  
 QY 707 FODEKMLSVLECLRTKMPREDLELETESLAMEDIKGTVVLLRLREEGVQVADIDFG 766  
 DB 522 LHHNMLPRVNSMLQVYLPARASLELEVTETGIMEDISTAOHLLSLRAGALLAIDDFG 581  
 QY 767 TGYSSLILKQLPRIHRLKIDKSEFVNDLNEGADTAIIQVYVIDLANGLELVAGIESSEA 826  
 DB 582 TGYSSLIVLKSLPLDKIKIDKSFVODLQEDDATTVAIIQIGKSLGQVIAEGEVTAE 641  
 QY 827 QLOLRQKMGCHLGGYFLTRPLPAAMMTYL 857  
 DB 642 QEAYIIAEGCNEGGYLYSKPLPARELTQYL 672

RESULT 11  
 ID P72843 PRELIMINARY; PRT; 840 AA.  
 AC P72843;  
 DT 01-FEB-1997 (TREMblrel. 02, Created)  
 DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE NITROGEN FIXATION POSITIVE ACTIVATOR PROTEIN.  
 GN NIFL OR SLR1305.  
 OS Synechocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
 RN NCBI\_TaxID=1148;  
 RP SEQUENCE FROM N.A.

RA MEDLINE=97061201; PubMed=8905231;  
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
 RA Miyajima N., Hiroseawa M., Sugitara M., Sasamoto S., Kimura T.,  
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naru K., Okumura S.,  
 RA Shimpou S., Takeuchi C., Wada T., Matenabe A., Yamada M., Yasuda M.,  
 RA Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the  
 RT entire genome and assignment of potential protein-coding regions.";  
 RT DNA Res. 3:109-136(1996).  
 CC -! SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER  
 CC REGULATORY COMPONENTS OF SENSOR TRANSDUCTION SYSTEMS.  
 DR EMBL: D90901; BAI16858.1; -.  
 DR HSP: P08402; I800.  
 DR InterPro: IPR001633; DUF2.  
 DR InterPro: IPR000160; DUF9.  
 DR InterPro: IPR001610; PAC.  
 DR InterPro: IPR000014; PAS.  
 DR InterPro: IPR000700; PAS-associat.  
 DR InterPro: IPR001789; Response-reg.  
 DR Pfam: PF00563; EAL; 1.  
 DR Pfam: PF00785; PAC; 1.  
 DR Pfam: PF00989; PAC; 2.  
 DR Pfam: PF00072; response-reg; 1.  
 DR SMART: SM00267; DUF1; 1.  
 DR SMART: SM00052; DUF2; 1.  
 DR SMART: SM00086; PAC; 2.  
 DR SMART: SM00091; PAS; 2.  
 DR SMART: SM00448; REC; 1.  
 KW Complete proteome; Phosphorylation; Sensory transduction.  
 SQ SEQUENCE 840 AA; 96226 MW; 5F56F3D367BD985 CRC64;

Query Match 12.1%; Score 801.5; DB 16; Length 840;  
 Best Local Similarity 28.2%; Pred. No. 4.8e-47;  
 Matches 234; Conservative 150; Mismatches 268; Indels 179; Gaps 24;

QY 140 LSSLT--VPVMDQQLMGIMAVHNSKPRFTEQEMETNALSKEYSLAITSQSLRQVQ 197  
 DB 86 LSAITDQSIYKGLGAVDYIH--KP--FQGE-----ITRQLQKLKH- 127  
 QY 198 QOVQALVQRLTETTYAQGDREPTWQVLALEYGAQVADGAVLY-IAD----- 245  
 DB 128 -QTNQALLQKNOLEBOIEKTAQTQALMOS-----EINFAVAFNSQDPFIYIGRNSGC 181  
 QY 246 -----LTG-SVAQHYQWNLRFDMNML-ETSIQOE----- 273  
 DB 182 IMDVQPRCPQFGGLTKQELVGVSRQGFVWVDEQROSLRDLMLQKDGLYKRENRETE 241  
 QY 274 -----LMRGQ-----SAMESPMA----- 288  
 DB 242 VYDQEKQVMTLSGEPFENRVDCLLFVMDIFERRKAQKQLTSQACQSPASIVIT 301  
 QY 289 -VQS--TWKRRPFTSVAPLRPTNCVPHGYTIGLEQSDMTAPRESLAENFQSFILVP 345  
 DB 302 DVQGNITVYVNPK-FEES-----GYKSAEVLGNPRILKSGNTQDYE-LMKKT 349  
 QY 346 LAADQWVGSLLILRKESLVKNHAGKRGIDRRNLRPLSFEAMEETOKLVPTWNRSRK 405  
 DB 350 LASGRNHGFEHNRKKNEL-YW-----ERASIFPISN----- 381  
 QY 406 LAQVASTOLYMAITQOFV-----TRLITQGTAYDPLTQLEPNIENRQLTALLDALYESG 460  
 DB 382 --QGIYVHYAAVKEIDIKKQQAELFNQAHYHDLTLPLRIKAKDLQQAISALRQK 439  
 QY 461 KMGVLYIAMDREKRNESFGHKKTGGDGLQEVADRNLNOKLSPLAAYSPLLSRMHGDDFT 520  
 DB 440 HIFGLMFDLNFKRVNDFTLGHADGDDQLVEVSESLRAL-----RQITGVARLGDEFLL 495  
 QY 521 LITQISDQEMIPCELRSLTFQEPFLOGQPIYLVTASMGISTAPYDGETAESLFAEIALTRAK 580  
 DB 496 ILDOVSHSRKLMAITQRLRLRVMPVNDGLLEFVHSGIGITVPPDDGFADVILLRNDT 555

[illegible]

RESULT	12		
P74400			
ID	P74400	PRELIMINARY;	PRT; 1578 AA.
AC	P74400;		
DT	01-FEB-1997 (TRMBLrel. 02, Created)		
DT	01-FEB-1997 (TRMBLrel. 02, Last sequence update)		
DT	01-OCT-2001 (TRMBLrel. 18, Last annotation update)		
DE	HYPOTHEMETICAL 178.5 KDA PROTEIN.		
GN	SLI0267.		
OS	Synechocystis sp. (strain PCC 6803).		
OC	Bacteria; Cyanobacteria; Chroococcales; Synechocystis.		
OX	NCBI_TaxId=1148;		
RP	[1]		
RF	SEQUENCE FROM N.A.		
RX	MEDLINE=97061201; Pubmed=8905231;		
RA	Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y., Miyajima N., Hirosewa M., Sugitani M., Sasamoto S., Kimura T., Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nario K., Okumura S., Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M., Tabata S.;		
RT	"Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.";		
RL	DNARES. 3:109-136(1996).		
DR	EMBL; D90914; BAA18497.1; -		
DR	InterPro: IPR001633; DUF2.		
DR	InterPro: IPR000160; DUF9.		
DR	InterPro: IPR003018; GAF.		
DR	InterPro: IPR001610; PAC.		
DR	InterPro: IPR000014; PAS.		
DR	InterPro: IPR000700; PAS-assoc_C.		
DR	Pfam; PF00563; EAL; 1.		
DR	Pfam; PF01590; GAF; 1.		
DR	Pfam; PF00990; GGDEF; 1.		
DR	Pfam; PF00785; PAC; 2.		
DR	Pfam; PF00989; PAS; 2.		
DR	SMART; SM00267; DUF1; 1.		
DR	SMART; SM00052; DUF2; 1.		
DR	SMART; SM00065; GAF; 1.		
DR	SMART; SM00086; PAC; 2.		
DR	SMART; SM00091; PAS; 2.		
DR	Hypothetical protein; Complete proteome.		
WQ	SEQUENCE 1578 AA; 178546 MW; BEI1604B0FBA556DF CRC64;		

Query Match	12.18;	Score 797;	DB 16;	Length 1578;
Best Local Similarity	29.88;	Pred. No. 2.8e+6;		
Matches 222;	Conservative 124;	Mismatches 241;	Indels 158;	Gaps

Db	952	GLFSSAAVPIRGGQVVAVLTFYSQAGEFRSEDOJLLEISDNLSFALDILQDDA--N	1009
QY	198	QOQVEBALVQLETTVAQYGDREPTEWQYALETVGAQVADGAVLYIAPDLTGSVAOHQWN	257
Db	1010	ORAN-----OLGNEERLRLALEAANOQ-----FYDN	1038
QY	258	LRFPMGMWLELSLMQELMRQPSAAAEPRMAAAGVSTMEKRPFTSVAPLRPNCPVHGTL	317
Db	1039	LQ-----TQO--AVVSPQ-----YQGL	1054
QY	318	G-----ELEORSDWIAPESLSANFQSFILVPLAADQWVGLILIRKESLVKMA	370
Db	1055	GYPDQYQVQEHQR--WL---ERHPQ-----DQ-----KREIYDQIQ	1089
QY	371	GKRGIDRRNLLPRLSEAWETO-----KLVPWNSEKRLAQVASTQLYMAIT	419
Db	1090	GK-----IPRAVACRQRTWAGQMKWLSLUGKITV--EMDSQGRPLMIG---ILTDVT	1137
QY	420	QOAFTRLITQOTA--YDPLTQDLPWITIFPNQTL-----ALDALYEGKMGVLYIANDRF	473
Db	1138	ERKQAEQIENLVAYPDLPALP-----NRRLDLRAENMLALAQSKHFGAIVLIDIDGF	1192
QY	474	KRINESFGHTGGGLQEOVADRLNOKTSLPAAYSPLSRWHDGPTILLTOSDNOEK--	531
Db	1193	KTLWDARHGDSGDRLOMAKRLADSLRD----SDTVAHLGDEFTVLLPPELAHDEELA	1248
QY	532	---IPLCERLSTFOEPEFFLOQGPILYLTASMGISTAPYUGFEASBLKFAEIALTRACQ	588
Db	1249	RLGIGVGEKIRQALAPFETLEAQVOQISGISTGLEPKENKVSDDLFEKADTAMYOQKA	1308
QY	589	GKNTYQFYRPDQSAFMLDRILTESDLRQALITNOEFLYHQPOVALDTSKLLGVEALYRWQ	648
Db	1309	GRDVCJLFESOMOLEVESREPLADRSALAEOKOFQVYLOPOVD--SHGVWIGAEALLRN	1367
QY	649	HPRGQVAPVETPLMEELGLNHLNQWLEFLACAHQHFFR--ETGRRLMAVNISARQ	706
Db	1368	HPRGGEFTPPNFTPLIAEILIGIGIDFVLEOVC--QYLAHLQDLGLTRIALNIVSPQ	1424
QY	707	FQDERKWTNSVLECKRGMPDELELEITESLAMEDIKGTVLLHRLREEGVOVALIDFG	766
Db	1425	FRQANPFAKEIKNLLACGVQVPRYLTEVTEBGLVEDBTHRALITMTELQTLGIHRSVDDG	1484
QY	767	TGYSLSILKQPLIHLKIDKSFVNDLNEGADTAIQYVIDLANGLMLFTVAIGISEA	826
Db	1485	TGYSLSLYLKRPLNELKIDRAEVQADQDNLNAAVLEALISVARTFMALIAVEGETEE	1544
QY	827	QLORLQKMGCHLQGYFLTPRLPAE	851
Db	1545	QVOFLAQDGNFYQGVYGGPRMD	1569

	ID	O9LAF1	PRELIMINARY;	PRT;	892 AA.
	AC	O9LAF1;			
DT	01-OCT-2000	(TRMBLrel. 15,	Created)		
DT	01-OCT-2000	(TRMBLrel. 15,	Last sequence update)		
DT	01-DEC-2001	(TRMBLrel. 19,	Last annotation update)		
DE	YKOW PROTEIN.				
GN	YKOW.				
OS	Bacillus cereus.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;				
CC	Bacillus/Staphylococcus group; Bacillus.				
OX	NCBI_TaxID=1396;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=ATCC 14579 TYPE STRAIN;				
RX	MEDLINE=20055637; Pubmed=10589720;				
RA	Oktad O., Gominet M., Purnelle B., Rose M., Iereclus D., Kolsto A.B.,				
RT	"Sequence analysis of three Bacillus cereus loci under PICR-regulated				
RL	genes encoding degradative enzymes and enterotoxin."				
EMBL	Microbiology 145:3129-3138(1999).				
	EMBL: AJ243712; CAB69812.1; -				



RC STRAIN-MAFE303099;  
 RX MEDLINE-21082930; PubMed-11214968;  
 RA Kaneo T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Watanabe A., Ideesawa K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.,  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 RT Mesorhizobium loti.";  
 RL DNA Res. 7:331-338(2000).  
 DR EMBL: AP003015; BAB54575.1;  
 DR InterPro: IPR001633; DUF2.  
 DR InterPro: IPR00160; DUF9.  
 DR Pfam: PF00563; EAL; 1.  
 DR Pfam: PF00990; GDEF; 1.  
 DR SMART: SM00267; DUF1; 1.  
 DR SMART: SM00052; DUF2; 1.  
 KW Plasmid: Complete proteome.  
 SQ SEQUENCE 653 AA; 71042 MW; 250B1D28BC5E61B8 CRC64;

Query Match 11.8%; Score 780.5; DB 16; Length 653;  
 Best Local Similarity 32.0%; Pred. No. 9.2e-46;  
 Matches 199; Conservative 109; Mismatches 228; Indels 85; Gaps 13;

QY 262 WGNMLET--LQELMRGQPS-AAEPMAAYOSTWEKPRPFTSVAPLPPTNCYPHGYTIG 318  
 DB 65 WGSPOOTKALWASYIAGAGFVGLRARSFSQV--KPR--SVSRRTQNLVRNAFLFG 119  
 QY 319 ELEORSDMIAAP-----ESLSAENFOSFLIPLVLAQQ-----WVGS 355  
 DB 120 -----TWGALVLPFGGAATNAQVYITCLSAQMTNGAASFTTITPAIAIATLPFVGS 174  
 QY 356 LILREKSLVKNWAGKRGIDRRNLLPRLSFEAMEETOKLPTWNRSEKRLAQNASTOLY 415  
 DB 175 AVAI-----VMEGAVNLPV-----ALIVSAITLF 201  
 QY 416 MAI-----TQGFRTLTQOT-ADPILTLPNMTIFNRQLTALLDALYEGKMGVLY 467  
 DB 202 RAVLAHASEFTOSFILTQTESENAIRRDVLTSLPNRPSFNERLENALVDARQEDHFAILL 261  
 QY 468 IAMDREKRLNESFGHKTGDLQEVADRNLNOKLSPLAASPLSRWHGCGFTILLTOISD 527  
 DB 262 PDLSNFDEVDHFGRAMADVLEMAARLRK-----TRESGIRARLEGGEFALIIARDIR 317  
 QY 528 NOEMIPLCERLLSTFOEPFLQGOPIYLTASMGISTAPYDGETAESLKFPAETALTRAC 587  
 DB 318 PQIGSLAKQIIDVMKAPPLIGREIYCRYSVIGIALAPTDGLDANQLRCVPTALHRAKT 377  
 QY 588 QGKNTYQFRPODSAPMLDRLTLESPLRQALTNQEFVLYFOPOVALDGTGKLGVEALVRW 647  
 DB 378 LAGSTQPFSSASDDDAARRNALERDLASALANDELMALFOPLDGSERIRGEFALLRW 437  
 QY 648 OHPRLGOVAPDVEIPLAEELGLINHGOWVLETTACATHOHFFETGRRLMAVNISAROF 707  
 DB 438 QHPTIGAIIPSEFIPIAETALIHITIGHWAKTACLAAYHWPRD-----LRVSNLSAVOL 493  
 QY 708 ODEKMLNSVLECIKRTGMPDELELEITESTLMEMEDIKGTVLLHRLREGVQVADIDFGT 767  
 DB 494 KDKALLDGIWALSTEGLEPKRLEVETESVLSIDPEDAISLLQSLSPSLSVVALDDFGT 553  
 QY 768 GYSSLSILKQPLIRHLKIKSFVNDLNGADTAIIQYVIDLANGLNLETVAGTISEAQ 827  
 DB 554 GFSSTLYLTKPLSRKIDRSFQDMADACAIIVRSVOLAHELRIEVTAGVETPEQ 613  
 QY 828 LORLOKMGCHLGQGYFLTRPL 848  
 DB 614 LDYLRVGCDEAQTILGKPV 634

THIS PAGE BLANK (USPTO)





XX This polypeptide comprises a phytochrome of the green alga  
CC *Mesostentium caldariorum*. The invention provides a new class of  
CC fluorescent protein adducts (phytofluors) that are generally  
CC suitable for use as fluorescent markers. They comprise a protein  
CC component (an apoprotein) and a bilin chromophore such as  
CC phycoerythrobilin. Preferred apoproteins are obtained from plants,  
CC e.g. oat (see AMW50144), from green algae, or from cyanobacteria (see  
CC AMW50143). Truncated apoproteins consisting of the N-terminal  
CC chromophore domain are especially preferred. Recombinant  
CC apoproteins assemble spontaneously with the bilin chromophore.  
CC Claimed compositions comprise a protein, glycoprotein, antibody or  
CC nucleic acid to be detected linked to the phytofluor. They are used  
CC in assays for detecting the other member of a specific binding pair,  
CC e.g. immunoassay of antigens, immuno-histochemical labelling, as  
CC nucleic acid probes for Southern blotting, for identification of  
CC manufactured products, also to detect protein-protein interactions,  
CC including studies on intracellular protein localisation and  
CC identification of transfected cells. The phytofluors make ideal  
CC fluorescent markers because they have a long wavelength absorption  
CC maximum and high molar absorption coefficient, and are stable to  
CC light and pH.  
XX  
SQ Sequence 1142 AA;

	Score	DB	Length
Query Match	76.78	785	19
Best Local Similarity	70.08	7.2e85	
Matches 147, Conservative		20	Mismatches 29
			Indels 14
			Gaps 1

```

Oy 1 KLRWASIRLOSJLGGDGLCALCDPVEDVQKLTGTDPRVMVYQFHEDDHGEVSIIRSDL 60
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 203 kLaaaaisrlsIsIpggdIglIcdaveevreIlgYdtrvmaYkfhedengeIaeItrsdI 262
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

Oy 61 EPYLGIAHPADIDIPQAAFELEKONRVMICCNATPYKVVSSEELKRPCLVNSTLEAPH 120
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 263 epyIglIhYpaddIpgaaftImkhrvrlIdcdseppYkvtIgdpmkmpIsIagstIlgvH 322
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

Oy 121 GCHOTYMMNMGSVASLALAIYVKGRDSS-----KIMGVLVGHCSPRYVP 166
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 323 gchagYmanngsvaSlvmaavlIndnsseogaIaagIlgkYkrIgmIgvchhseprYvpf 382
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

Oy 167 PLRYACETLMAFGIQLQLOMELQLASQLAEK 196
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 383 pLrsaceIImgvIglqlmmeveIsSqIreK 412
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

RESULT	2
AAG62367	
ID	AAG62367 standard; Protein; 1128 AA.

DT	29-AUG-2001	(first entry)
XX		
DE	AP4 related amino acid sequence.	
XX		
KW	Phenotype switch molecule; phenotype-related gene battery; AP4.	
KW	gene localisation.	
XX		
OS	Unidentified.	
XX		
PN	WO200138515-A1.	
XX		
PD	31-MAY-2001.	
XX		
PF	17-NOV-2000; 2000WO-CN00427.	
XX		
PR	19-NOV-1999; 99CN-0121466.	
XX		
PA	(BIAN/) BIAN X.	
XX		
PI	BIAN X;	

XX WP1: 2001-367684/38.  
 DR  
 XX Isolating phenotype switch molecules and phenotype-related gene  
 PT batteries from complex genomes of higher animals and plants, useful  
 PR e.g. in gene localization and classification analysis -  
 PX  
 XX  
 PS Example 8; Page 30-31, 35pp; Chinese.

This invention relates to a method for isolating phenotype switch molecules and phenotype-related gene batteries from complex genomes of higher animals and plants. The method is useful in gene localisation and classification analysis, studying gene development networks and function networks, and designing drugs based on regulatory sequences of the phenotypes for disease treatment. The present sequence represents an AP4 related protein, which is used in an example illustrating the use of the method of the invention.

Sequence 1128 AA;

Query Match	65.1%	Score 666.5;	DB 22;	Length 1128;
Best Local Similarity	59.7%;	Pred No. 1.3e-70;		
Matches 126;	Conservative 26;	Mismatches 44;	Indels 15;	Gaps 1;

```

OY 1 KLNRAISRLOSLEFGDGLACDPVWEDYQKLTCTYDWMVYVOFHEDDHGVSEIRSDL 60
Db 200 kllaakatsktslpgssmvevcltvmxkvefdl byddvrmayk ftheddhgevfaeltkpgl 259
OY 61 EPYGLHYPADPIDOARFELKONRVRVICCNATPKVVOSEELKRPCLLVNSTLEAPH 120
Db 260 epylgllhpadldpgaarfltmkxkvzmicdctrarskvtlaeealprfdlslsgsaltraph 319
OY 121 GCHTOYMANMGSVASLALAIYV-----KGDSSKTLGWLGVGHHCSPRYVP 165
Db 320 schlgymenmnstslslymavvvneeeddaeeseqpqdqkklkwlglvlchhespryvp 379
OY 166 FPLRACFELQWAFGLQLOMLQTLASQLAEK 196
Db 380 fplryaceflaqvafvahnvrefelekqlire 410

```

RESULT	3
AAW50144	
ID	AAW50144 standard; protein; 1129 AA

DT	28-AUG-1998	(first entry)
XX		
DE	Oat phytochrome A apoprotein.	
XX		
KW	Phytofluor; fluorescent label; phytochrome A; oat.	
XX		
OS	Avena sativa.	
XX		
PN	MO9805944-A1.	
XX		
PD	12-FEB-1998.	
XX		
PF	01-AUG-1997; 97MO-US13529.	
XX		
PR	02-AUG-1996; 96US-0023217.	
XX		
PA	(REGC ) UNIV CALIFORNIA.	
XX		
PI	Lagararias JC, Murphy JT;	
XX		
DR	WPI; 1998-145711/13.	
XX		
PT	Adducts of apoprotein polypeptide and chromophore as label,	
PT	particularly for bio-molecules - used as fluorescent markers in	
PT	immunoassays, nucleic acid hybridisation, detecting protein-protein	

```
PT interaction etc., are stable with high molar absorption
```

```
XX Example 1; Page 63; 87pp; English.
```

```
PS  
CC This polypeptide comprises phytochrome A (PhyA) of oat. A method  
CC for the efficient purification of recombinant PhyA from  
CC Saccharomyces cerevisiae using a C-terminal epitope tag is  
CC provided. The invention provides a new class of fluorescent  
CC protein adducts (designated phytofluors) that are generally suitable  
CC for use as fluorescent markers. They comprise a protein component  
CC (an apoprotein) and a bilin chromophore such as phycoerythrobilin.  
CC Preferred apoproteins are obtained from plants, e.g. oats, from  
CC green algae, e.g. Mesocostium caldariorum (see AA050145), or from  
CC cyanobacteria such as Synechocystis (see AA050143). Truncated  
CC apoproteins consisting of the N-terminal chromophore domain are  
CC especially preferred. Recombinant apoproteins assembled  
CC spontaneously with the bilin chromophore. Claimed compositions  
CC comprise a protein, glycoprotein, antibody or nucleic acid to be  
CC detected linked to the fluorescent adduct. They are used in assays  
CC for detecting the other member of a specific binding pair, e.g.  
CC immunassay of antigens, immuno-histochemical labelling, as nucleic  
CC acid probes for Southern blotting, for identification of manufactured  
CC products, also to detect protein-protein interactions, including  
CC studies on intracellular protein localisation and identification of  
CC transfected cells. The phytofluors make ideal fluorescent markers  
CC because they have a long wavelength absorption maximum and high  
CC molar absorpton coefficient, and are stable to light and pH.
```

```
SO Sequence      1129 AA;
```

```
Query Match          64.8%; Score 663.5; DB 19; Length 1129;  
Best Local Similarity   59.7%; Pred. No. 3e-70;  
Matches    126; Conservative     24; Mismatches   46; Indels     15; Gaps       1.
```

```
OY      1 KLAVALAIRSLQSPGCDIGALCITYVEDVQRITGDRAWNVYGFHEDDHEEVSSEIRSDL 60  
Db      |||:::|||||::|||::|||::|||::|||::|||::|||::|||::|||::|||:  
OY      201 KAAAKAIKSIQSIPGSMEVCNIVKEVFIDLTGYDRMAVKIFHNDDHGVEFSLEIKPIL 260  
Db      |KKAAAKIksiqspgsmevcnivkvefidltgydrmavkfifhnddhgvefsleikpil  
  
OY      61 EPLYIAHPADVIDPAARFLFKONRVRMTCDNATPVKVVOSEEELRPCLINSTLRAPH 120  
Db      261 epylighpadidipaarallfmkkvrmncdcratsirkieaalpfdislgsalaraph 320  
  
OY      121 GCCHTOYMANNMGSVASLALAIVWKKD-----SSKLGLGVGHHCSDRYVP 165  
Db      |||||||::|||::|||::|||::|||::|||::|||::|||::|||::|||  
        schligymennmsiaslvmavvnneneedeaeeseqpaaqqakkkkiwllgvchespryp 380  
        fplyacefiagvfavnrvrefekqlrek 411
```

```
OY      166 FPLRAYCEFIHQAFGILOMETQLALSOLAALK 196  
Db      |||||||::::::::::|||::|||::|||::|||::|||::|||::|||::|||  
        flrayacefiahqafgiolometqlalsolaalk 196
```

```
Dd      381 flryacefiagvfavnrvrefekqlrek 411
```

```
RESULT  4.  
AADB26592  
XZ AAB26592 standard; protein; 748 AA.
```

```
XZ AAB26592;  
AC  
DT 01-FEB-2001 (first entry)  
DE Synchocystis sp phytochrome-related gene CphI.  
KW Phytochrome; fluorescent apophytochrome; bilin; Cph.  
OS Synechocystis sp.  
XX MO200056355-A1.  
XX PD 28-SEP-2000.  
XX PF 14-MAR-2000; 2000OWO-US06607.  
XX PR 19-MAR-1999; 99US-0272809.
```

[illegible]

```

FT      /note= "histidine kinase transmitter module
FT      conserved motif"
XX      WO09805944-1.
XX      12-FEB-1998.
XX      01-AUG-1997; 97WO-US13529.
XX      02-AUG-1996; 96US-0023217.
XX      (REGC ) UNTV CALIFORNIA.
XX      Lagariae JC, Murphy JT;
XX      WPI, 1998-145711/13.
XX
XX      Products of apoprotein polypeptide and chromophore as label,
XX      particularly for bio-molecules - used as fluorescent markers in
XX      immunoassays, nucleic acid hybridisation, detecting protein-protein
XX      interaction etc., are stable with high molar absorption
XX
XX      Example 4; Fig 10B; 87pp; English.
XX
XX      This polypeptide comprises cyanobacterial phytochrome 1 (Cph1) of
XX      Synechocystis sp. PCC6803. Its amino acid sequence was deduced
XX      from locus slr0473 genomic DNA. Expression of the 748-residue
XX      polypeptide in E. coli and incubation with phycoerythrobilin yields
XX      an adduct with a red, far-red photoreversible phytochrome
XX      signature. The invention provides a new class of fluorescent
XX      protein adducts (designated phytofluors) that are generally suitable
XX      for use as fluorescent markers. They comprise a protein component
XX      (an apoprotein) and a bilin chromophore such as phycoerythrobilin.
XX      Preferred apoproteins are obtained from plants, e.g. oats (see
XX      AAW50144), from green algae, e.g. Mesotetium caldariorum (see
XX      AAW50145), or cyanobacteria such as Synechocystis. Truncated
XX      apoproteins consisting of the N-terminal chromophore domain are
XX      especially preferred. Recombinant apoproteins assemble
XX      spontaneously with the bilin chromophore. Claimed compositions
XX      comprise a protein, glycoprotein, antibody or nucleic acid to be
XX      detected linked to the fluorescent adduct. They are used in assays
XX      for detecting the other member of a specific binding pair, e.g.
XX      immunoassay of antigens, immuno-histochemical labelling, as nucleic
XX      acid probes for Southern blotting, for identification of manufactured
XX      products, also to detect protein-protein interactions, including
XX      studies on intracellular protein localisation and identification of
XX      transfected cells. The phytofluors make ideal fluorescent markers
XX      because they have a long wavelength absorption maximum and high
XX      molar absorption coefficient, and are stable to light and pH.
XX
XX      Sequence 748 AA:
XX
XX      Query Match 28.5%; Score 292; DB 19; Length 748;
XX      Best Local Similarity 36.8%; Pred. No. 8.5e-26;
XX      Matches 64; Conservative 38; Mismatches 64; Indels 8; Gaps
XX
XX      2 LAVRAISRLOSPEGDIGALCDTVVEDVQRLTGYDRWVYQFREDHGEVSEIRSDLE 61
XX      :||::||::: ||::||::||::||::||::||::||::||::||::||::||::||::||
XX      Db 139 manaaInrlrTg--qanlrfdfgydVveevrmifqgvmlYtrdennhcvldkrdde 196
XX
XX      62 PVLGHPPTDIPQARFLFKONRVMIDCN--ATPVVQVSEELKRLCLVNSTLRAP 119
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      Db 197 pIlqIlhpesdIpqprtrIfImprIvPdyvgyavprIpravnpsatnradvltlesIlTs 256
XX
XX      120 HGGCHQDYANMMSVASIALAIIVKGGDSKTLGLVGNHCSPPVFPFLRYACE 173
XX      :|| ||::||| ||::||::||::||::||::||::||::||::||::||::||::||::||
XX      Db 257 yncHtLykImngvgasttsIlkdg----hlwglIachnqtrkpvInelrkege 306
XX
XX      RESULT 6
XX      AAB26600 standard; peptide: 212 AA.

```

```

XX AAB26600;
AC
XX
XX 01-FEB-2001 (first entry)
DT
XX
XX Synechocystis sp phytochrome-related gene Cph2-N197 peptide.
DE
XX
XX Phytochrome; fluorescent apophytochrome; bilin; Cph.
KM
XX
XX Synechocystis sp.
OS
XX
XX WO200056355-A1.
PN
XX
XX 28-SEP-2000.
PD
XX
XX 14-MAR-2000; 2000MO-US06607.
PF
XX
XX 19-MAR-1999; 9905-0272809.
PR
XX
XX (REGC ) UNIV CALIFORNIA.
PA
XX
XX Lagarias JC;
PI
XX
XX WPI; 2000-602195/57.
DR
XX
XX Fluorescent apophytochrome-bilin conjugates, e.g. Cph2 (derived from
PT Synechocystis species) and phycoerythrobilin conjugate, useful as
PT fluorescent markers for biological research -
XX
XX Claim 5; Fig 3; 52pp; English.
PS
XX
XX The present invention is related to fluorescent apophytochrome-bilin
CC conjugates, known as phytofluors. An apoprotein known as Cph2 from
CC Synechocystis species is used as the apophytochrome and the bilin is
CC preferably phycoerythrobilin. The phytofluors are useful as fluorescent
CC markers for biological research. The phytofluors have a long wavelength
CC absorption maxima, a high molar absorption coefficient and the
CC recombinant apoproteins can spontaneously assemble with a variety of
CC bilin chromophore precursors. The present sequence is a phytochrome
CC related peptide from Synechocystis sp.
XX
XX
XX Sequence 212 AA:
SQ

```

```
OS Synechocystis sp.
XX
XX WO200056355-A1.
XX
XX 28-SEP-2000.
XX
XX 14-MAR-2000; 2000WO-US06607.
XX
XX 19-MAR-1999; 99US-0272809.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Lagarias JC;
XX
XX WPI: 2000-602195/57.
XX
XX Fluorescent apophytochrome-bilin conjugates, e.g. Cph2 (derived from
XX Synechocystis species) and phycoerythrobilin conjugate, useful as
XX fluorescent markers for biological research -
XX
XX Example 1; Page 45-46; 52pp; English.
XX
XX The present invention is related to fluorescent apophytochrome-bilin
XX conjugates, known as phytofluors. An apoprotein known as Cph2 from
XX Synechocystis species is used as the apophytochrome and the bilin is
XX preferably phycoerythrobilin. The phytofluors are useful as fluorescent
XX markers for biological research. The phytofluors have a long wavelength
XX absorption maxima, a high molar absorption coefficient and the
XX recombinant apoproteins can spontaneously assemble with a variety of
XX bilin chromophore precursors. The present sequence is a phytochrome
XX related protein from Synechocystis sp.
XX
XX Sequence 1276 AA:
SQ
Query Match 18.6%; Score 190; DB 21; Length 1276;
Best Local Similarity 31.5%; Pred. No. 3.5e-13;
Matches 51; Conservative 24; Mismatches 67; Indels 20; Gaps 5;
OY 5 RAISRLQSLPGDIGALCDTVVEDVQRLTGXYDRVWVYGFHEDHGEVSE-IRRSDLPEY 63
DB 20 ralfrel-----qviveearlflgydrvklykfadsgevaavnaalpsl 70
OY 64 LGLHYPATDIPQAARFLFKQNRVRMICDCNATPVKVVQSEELK--RPLCLVNSTLRAPH 120
DB 71 LglHfpyedlppqareelngqrxmlavdvahnrkk---shelsgrlsptlehsnglytlvd 127
OY 121 GCHROYMANMGSVASIALAIVVKGKDSKLMGIVVGHHCSPR 162
DB 128 schiyllamgvlsiltvpm-----gdqqlwglmaavhnskpr 165
RESULT 8
AAB26597
ID AAB26597 standard; protein; 844 AA.
AC AAB26597;
XX
XX 01-FEB-2001 (first entry)
XX
XX Synechocystis sp phytochrome-related gene Cph6.
XX
XX Phytochrome; fluorescent apophytochrome; bilin; Cph.
XX
XX Synechocystis sp.
XX
XX WO200056355-A1.
XX
XX 28-SEP-2000.
XX
XX 14-MAR-2000; 2000WO-US06607.
XX
XX 19-MAR-1999; 99US-0272809.
XX
```

```
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Lagarias JC;
XX
XX WPI: 2000-602195/57.
XX
XX Fluorescent apophytochrome-bilin conjugates, e.g. Cph2 (derived from
XX Synechocystis species) and phycoerythrobilin conjugate, useful as
XX fluorescent markers for biological research -
XX
XX Example 1; Page 48-49; 52pp; English.
XX
XX The present invention is related to fluorescent apophytochrome-bilin
XX conjugates, known as phytofluors. An apoprotein known as Cph2 from
XX Synechocystis species is used as the apophytochrome and the bilin is
XX preferably phycoerythrobilin. The phytofluors are useful as fluorescent
XX markers for biological research. The phytofluors have a long wavelength
XX absorption maxima, a high molar absorption coefficient and the
XX recombinant apoproteins can spontaneously assemble with a variety of
XX bilin chromophore precursors. The present sequence is a phytochrome
XX related protein from Synechocystis sp.
XX
XX Sequence 844 AA:
SQ
Query Match 12.0%; Score 123; DB 21; Length 844;
Best Local Similarity 22.2%; Pred. No. 2.2e-05;
Matches 46; Conservative 39; Mismatches 72; Indels 50; Gaps 7;
OY 11 QSLPGDIGALCDTVVEDVQRLTGXYDRVWVYGFHEDH--GEVYSEIRSDLEPYLGIH 67
DB 472 gsl---dlplfnvtvgeirfgleadrviqfqsdsfsgnivaesvlapfkplinsa 528
OY 68 YPATDIPQAARFLFKQNRVRMICDCNATPVKVVQSEELKRPCLVNSTLRAPHGCHTOYM 127
DB 529 leelfsmnagryqgrlqviedlh-----qshlrq---chidfl 566
OY 128 ANMGSVASIALAIVVKGKDSKLMGLVVGHCSPRYV-----PPPLRYA 171
DB 567 arlqyranlvlpil-----ndallkwllclchgdsarvvegteidllkqdtngfelaigqa 622
OY 172 C--EFLMQAFGLQLOMELQLASQLAEK 196
DB 623 tlyeqaqgelasknqlfvqltnelegk 649
RESULT 9
AAB26594
ID AAB26594 standard; protein; 481 AA.
AC AAB26594;
XX
XX 01-FEB-2001 (first entry)
XX
XX Synechocystis sp phytochrome-related gene Cph3.
XX
XX Phytochrome; fluorescent apophytochrome; bilin; Cph.
XX
XX Synechocystis sp.
XX
XX WO200056355-A1.
XX
XX 28-SEP-2000.
XX
XX 14-MAR-2000; 2000WO-US06607.
XX
XX 19-MAR-1999; 99US-0272809.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Lagarias JC;
XX
```

DR WPI; 2000-602195/57.  
XX  
XX Fluorescent apophytochrome-bilin conjugates, e.g. Cph2 (derived from  
PT Synechocystis species) and phycoerythrobilin conjugate, useful as  
PT fluorescent markers for biological research -  
XX  
XX Example 1; Page 46; 52pp; English.  
PS  
XX The present invention is related to fluorescent apophytochrome-bilin  
CC conjugates, known as phytofluors. An apoprotein known as Cph2 from  
CC Synechocystis species is used as the apophytochrome and the bilin is  
CC preferably phycoerythrobilin. The phytofluors are useful as fluorescent  
CC markers for biological research. The phytofluors have a long wavelength  
CC absorption maxima, a high molar absorption coefficient and the  
CC recombinant apoproteins can spontaneously assemble with a variety of  
CC bilin chromophore precursors. The present sequence is a phytochrome  
CC related protein from Synechocystis sp.  
XX  
XX Sequence 481 AA;  
SQ

Query Match 11.5%; Score 118; DB 21; Length 481;  
Best Local Similarity 20.3%; Pred. No. 4e-05; Mismatches 70; Indels 46; Gaps 4;  
Matches 39; Conservative 37; Mismatches 70; Indels 46; Gaps 4;  
QY 11 QSLPGDIGALCDTVEDVQRLTGYDRVMYQFHEDDGEVVSSEIRSDLEPYLGLHYPA 70  
DB 55 qsl---nlevlnttvaevkcllgvdryllyrtwdqgtgsaitesvanapsilgrtfsd 111  
QY 71 TDIQARFLFKQNRVRNICCNATPVKVVQSEELKRLCLVNSTLRPHGCHQYMANM 130  
DB 112 evfveyhdaytkgvtainididddei-----cladfvkqf 149  
QY 131 GSVASLALAIYVKKKDS-----KLMGLVGHHCSPRYVPPFLRYACEFLMQAFGLQ 182  
DB 150 gvkaklvvpilqhnraasldesefpylwgllthgca-----ftwpwgwe 196  
QY 183 LQMLQLASQLA 194  
DB 197 velmkqlangya 208

RESULT 10  
AAB26599  
ID AAB26599 standard; protein; 750 AA.  
XX  
XX AAB26599;  
AC  
XX 01-FEB-2001 (first entry)  
DT  
XX  
XX Synechocystis sp phytochrome-related gene Cph8.  
DE  
XX Phytochrome; fluorescent apophytochrome; bilin; Cph.  
KM  
XX Synechocystis sp.  
OS  
XX WO200056355-A1.  
PN  
XX 28-SEP-2000.  
PD  
XX  
XX 14-MAR-2000; 2000WO-US06607.  
PF  
XX  
XX 19-MAR-1999; 99US-0272809.  
PR  
XX (REGC ) UNIV CALIFORNIA.  
PA  
XX  
XX Lagarias JC;  
PI  
XX WPI; 2000-602195/57.  
DR  
XX  
XX Fluorescent apophytochrome-bilin conjugates, e.g. Cph2 (derived from  
PT Synechocystis species) and phycoerythrobilin conjugate, useful as  
PT fluorescent markers for biological research -  
PT

XX  
XX Example 1; Page 50; 52pp; English.  
PS  
XX  
XX The present invention is related to fluorescent apophytochrome-bilin  
CC conjugates, known as phytofluors. An apoprotein known as Cph2 from  
CC Synechocystis species is used as the apophytochrome and the bilin is  
CC preferably phycoerythrobilin. The phytofluors are useful as fluorescent  
CC markers for biological research. The phytofluors have a long wavelength  
CC absorption maxima, a high molar absorption coefficient and the  
CC recombinant apoproteins can spontaneously assemble with a variety of  
CC bilin chromophore precursors. The present sequence is a phytochrome  
CC related protein from Synechocystis sp.  
XX  
XX Sequence 750 AA;  
SQ

Query Match 11.4%; Score 117; DB 21; Length 750;  
Best Local Similarity 26.5%; Pred. No. 0.0001;  
Matches 49; Conservative 26; Mismatches 60; Indels 50; Gaps 11;  
QY 1 KLAVALAISRL-QSLPGDIGALCDTVEDVQRLTGYDRVMYQF----- 44  
DB 160 klsgvtaqirgsl---dlseilmavtavgkflfvdvlyvqfhygspsltpleengip 216  
QY 45 ---EDDHGEVVSSEIRSDLEPYLGLHYPAVDIPQARFLFKQNRVRNICCNATPVK 99  
DB 217 aprprqygvevtyearspeldtmjmtendcsqv--fayeqkyk-----gavav 268  
QY 100 VQSE-ELKRPICLVNSTLRPHGCHQYMANMGSVALAIYVKKKDSKLMGLVQHN 158  
DB 269 sdlenhysssyelv-----gllqryqr-----akivapilveg---qlwglilhqg 312  
QY 159 C-SPR 162  
DB 313 chpr 317

RESULT 11  
AAB26595  
ID AAB26595 standard; protein; 1371 AA.  
XX  
XX AAB26595;  
AC  
XX 01-FEB-2001 (first entry)  
DT  
XX  
XX Synechocystis sp phytochrome-related gene Cph4.  
DE  
XX  
XX Phytochrome; fluorescent apophytochrome; bilin; Cph.  
KM  
XX Synechocystis sp.  
OS  
XX WO200056355-A1.  
PN  
XX 28-SEP-2000.  
PD  
XX  
XX 14-MAR-2000; 2000WO-US06607.  
PF  
XX  
XX 19-MAR-1999; 99US-0272809.  
PR  
XX (REGC ) UNIV CALIFORNIA.  
PA  
XX  
XX Lagarias JC;  
PI  
XX WPI; 2000-602195/57.  
DR  
XX  
XX Fluorescent apophytochrome-bilin conjugates, e.g. Cph2 (derived from  
PT Synechocystis species) and phycoerythrobilin conjugate, useful as  
PT fluorescent markers for biological research -  
XX  
XX Example 1; Page 47; 52pp; English.  
XX  
XX The present invention is related to fluorescent apophytochrome-bilin  
CC conjugates, known as phytofluors. An apoprotein known as Cph2 from  
CC









**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2002, 18:55:44 ; Search time 25.57 seconds  
(without alignments)  
187.228 Million cell updates/sec

Title: US-09-272-809-9  
Perfect score: 1024  
Sequence: 1 KIAVRAISRLQSLPGSDIGA.....QAFGLQLQMLQASLAERK 196

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/CTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfilest.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the total being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	785	76.7	611	3 US-08-904-871-4	Sequence 4, Appli
2	785	76.7	1142	3 US-08-904-871-12	Sequence 12, Appli
3	733	71.6	554	3 US-08-904-871-1	Sequence 1, Appli
4	672	65.6	600	3 US-08-904-871-3	Sequence 3, Appli
5	663.5	64.8	600	3 US-08-904-871-2	Sequence 2, Appli
6	663.5	64.8	1129	3 US-08-904-871-11	Sequence 11, Appli
7	388	37.9	748	3 US-08-904-871-5	Sequence 5, Appli
8	388	37.9	748	3 US-08-904-871-6	Sequence 6, Appli
9	388	37.9	748	3 US-08-904-871-13	Sequence 13, Appli
10	72.5	7.1	353	1 US-08-197-792-43	Sequence 43, Appli
11	72.5	7.1	353	1 US-08-459-850-43	Sequence 43, Appli
12	72.5	7.1	353	1 US-08-459-214-43	Sequence 43, Appli
13	72.5	7.0	936	5 PCT-US93-05944-2	Sequence 6, Appli
14	71.5	7.0	5588	4 US-09-036-987A-6	Sequence 6, Appli
15	71.5	7.0	5588	4 US-09-370-700-6	Sequence 6, Appli
16	69	6.7	629	1 US-08-250-740-33	Sequence 33, Appli
17	69	6.7	629	1 US-07-695-472B-2	Sequence 2, Appli
18	69	6.7	1118	4 US-09-379-523-3	Sequence 3, Appli
19	68.5	6.6	2584	3 US-08-936-135-4	Sequence 4, Appli
20	67.5	6.6	308	4 US-09-347-803-8	Sequence 8, Appli
21	67.5	6.6	2588	3 US-08-936-135-2	Sequence 2, Appli
22	67	6.5	4928	4 US-09-036-987A-5	Sequence 5, Appli
23	67	6.5	4928	4 US-09-370-700-5	Sequence 5, Appli
24	66.5	6.5	349	1 US-08-197-792-33	Sequence 33, Appli
25	66.5	6.5	349	1 US-08-459-850-33	Sequence 33, Appli
26	66.5	6.5	349	1 US-08-459-214-33	Sequence 33, Appli
27	66.5	6.5	414	2 US-08-845-161A-2	Sequence 2, Appli

28	66.5	6.5	414	4 US-09-270-751-2	Sequence 2, Appli
29	65.5	6.4	178	4 US-09-147-928-4	Sequence 4, Appli
30	65.5	6.4	413	4 US-09-147-928-2	Sequence 2, Appli
31	65	6.3	206	4 US-09-311-311C-22	Sequence 22, Appli
32	65	6.3	313	4 US-09-347-803-25	Sequence 25, Appli
33	65	6.3	582	2 US-08-989-386-1	Sequence 1, Appli
34	64.5	6.3	399	2 US-08-742-621-4	Sequence 4, Appli
35	64.5	6.3	399	2 US-08-750-134A-5	Sequence 5, Appli
36	64.5	6.3	399	4 US-09-363-745-5	Sequence 5, Appli
37	64.5	6.3	993	1 US-08-468-557-2	Sequence 2, Appli
38	64	6.2	2548	4 US-09-172-422-1	Sequence 1, Appli
39	63.5	6.2	151	1 US-08-332-576-3	Sequence 3, Appli
40	63.5	6.2	151	5 PCT-US95-13672-3	Sequence 3, Appli
41	63.5	6.2	323	2 US-08-540-804-18	Sequence 18, Appli
42	63.5	6.2	323	2 US-08-218-265-18	Sequence 18, Appli
43	63.5	6.2	323	4 US-08-521-872-18	Sequence 18, Appli
44	63.5	6.2	323	3 US-08-590-399-18	Sequence 18, Appli
45	63.5	6.2	394	2 US-08-555-568B-17	Sequence 17, Appli

## ALIGNMENTS

```
RESULT 1
US-08-904-871-4
; Sequence 4, Application US/08904871
; Patent No. 6046014
; GENERAL INFORMATION:
; APPLICANT: Lagarias, John C
; APPLICANT: Murphy, John T
; TITLE OF INVENTION: PHOTOLUORS AS FLUORESCENT LABELS
; FILE REFERENCE: 2500.134USO UC OTT Lagarias Patent
; CURRENT APPLICATION NUMBER: US/08/904,871
; CURRENT FILING DATE: 1997-08-01
; EARLIER APPLICATION NUMBER: 60/023,217
; EARLIER FILING DATE: 1996-08-02
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO: 4
; LENGTH: 611
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:: Sequence
; OTHER INFORMATION: alignment Fig. 6 Mcphylb
US-08-904-871-4

Query Match 76.7%; Score 785; DB 3; Length 611;
Best Local Similarity 70.0%; Pred. No. 1.4e-90;
Matches 147; Conservative 20; Mismatches 29; Indels 14; Gaps 1;

QY 1 KIAVRAISRLQSLPGSDIGALCDTVVEDVORLTGYRVWVVOFHEDHGEVSEIRSDL 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 203 KIAAARISRLQSLPGSDILLCDAVVEVRELTGDRVAVVVFHDEHGEVIAEIRSDI 262
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 EBYLGLHYPATDIPQAAEFLFKQNRVMIICDNATPEPVYQSEELKRPCLVNSTLRAPH 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 263 EBYLGLHYPATDIPQAAEFLFKQNRVMIICDNATPEPVYQSEELKRPCLVNSTLRAPH 322
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 GCHTQYMANMGSVASIALAIVVKGKDS-----KIKLVVGHHCSPRYVP 166
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 323 GCHTQYMANMGSVASIALAIVVKGKDS-----KIKLVVGHHCSPRYVP 382
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 167 PLRYACEFLMAQFGLQMLQMLQASLAERK 196
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 383 PLRYACEFLMAQFGLQMLQMLQASLAERK 412
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 2
US-08-904-871-12
; Sequence 12, Application US/08904871
; Patent No. 6046014
```

```

; GENERAL INFORMATION:
; APPLICANT: Lagarias, John C
; APPLICANT: Murphy, John T
; TITLE OF INVENTION: PHYTOFLUORS AS FLUORESCENT LABELS
; FILE REFERENCE: 2500.134USO UC OTT Lagarias Patent
; CURRENT APPLICATION NUMBER: US/08/904,871
; CURRENT FILING DATE: 1997-08-01
; EARLIER APPLICATION NUMBER: 60/023,217
; EARLIER FILING DATE: 1996-08-02
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO: 12
; LENGTH: 1142
; TYPE: PRT
; ORGANISM: Alga (Mesotaelinum)
US-08-904-871-12

```

```

Query Match          76.7%; Score 785; DB 3; Length 1142;
Best Local Similarity 70.0%; Pred. No. 3,7e-90;
Matches 147; Conservative 20; Mismatches 29; Indels 14; Gaps 1;

```

```

Qy 1 KLAVAISRLQSLPGDIGALCDTVEEDVQRLTGYDRVAVVYQFHEDDGEVVSSEIRSDL 60
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 203 KLAARAIISRLQSLPGDIGALCDTVEEDVQRLTGYDRVAVVYQFHEDDGEVVSSEIRSDL 262
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 61 EPLYGLHYPATDIPQARFLFKQNRVIMICDNCATPPVKVYQSEELKRPCLVNSTLRAPH 120
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 263 EPLYGLHYPATDIPQARFLFKQNRVIMICDNCATPPVKVYQSEELKRPCLVNSTLRAPH 322
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 121 GCHTOYMANMGVSASIALAIYVKGDS-----KLMGLVVGHHCSPPRYVPF 166
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 333 GCHTOYMANMGVSASIALAIYVKGDS-----KLMGLVVGHHCSPPRYVPF 382
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 167 PLRYACEFLMOAFGLQLOMELQASOLAER 196
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 383 PLRYACEFLMOAFGLQLOMELQASOLAER 412
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

RESULT 3
US-08-904-871-1
; Sequence 1, Application US/08904871
; Patent No. 6046014
; GENERAL INFORMATION:
; APPLICANT: Lagarias, John C
; APPLICANT: Murphy, John T
; TITLE OF INVENTION: PHYTOFLUORS AS FLUORESCENT LABELS
; FILE REFERENCE: 2500.134USO UC OTT Lagarias Patent
; CURRENT APPLICATION NUMBER: US/08/904,871
; CURRENT FILING DATE: 1997-08-01
; EARLIER APPLICATION NUMBER: 60/023,217
; EARLIER FILING DATE: 1996-08-02
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO: 1
; LENGTH: 554
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:: Sequence
; OTHER INFORMATION: alignment Fig. 6 Consensus
US-08-904-871-1

```

```

Query Match          71.6%; Score 733; DB 3; Length 554;
Best Local Similarity 72.5%; Pred. No. 4,8e-84;
Matches 145; Conservative 20; Mismatches 27; Indels 8; Gaps 5;

```

```

Qy 1 KLAVAISRLQSLPGDIGALCDTVEEDVQRLTGYDRVAVVYQFHEDDGEVVSSEIRSDL 60
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 KLAARAIISRLQSLPGDIGALCDTVEEDVQRLTGYDRVAVVYQFHEDDGEVVSSEIRSDL 238
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 61 EPLYGLHYPATDIPQARFLFKQNRVIMICDNCATPPVKVYQSEELKRPCLVNSTLRAPH 120
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

Db 239 EPLYGLHYPATDIPQARFLFKQNRVIMICDNCATPPVKVYQSEELKRPCLVNSTLRAPH 296
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 121 GCHTOYMANMGVSASIALAIYVKGDS-----SKLMGLVVGHHCSPPRYVPFPLRYACEFLM 176
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 297 GCHTOYMANMGVSASIALAIYVKGDS-----SKLMGLVVGHHCSPPRYVPFPLRYACEFLM 356
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 177 QAFGLQLOMELQASOLAER 196
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 357 QAFGLQLOMELQASOLAER 376
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

RESULT 4
US-08-904-871-3
; Sequence 3, Application US/08904871
; Patent No. 6046014
; GENERAL INFORMATION:
; APPLICANT: Lagarias, John C
; APPLICANT: Murphy, John T
; TITLE OF INVENTION: PHYTOFLUORS AS FLUORESCENT LABELS
; FILE REFERENCE: 2500.134USO UC OTT Lagarias Patent
; CURRENT APPLICATION NUMBER: US/08/904,871
; CURRENT FILING DATE: 1997-08-01
; EARLIER APPLICATION NUMBER: 60/023,217
; EARLIER FILING DATE: 1996-08-02
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO: 3
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:: Sequence
; OTHER INFORMATION: alignment Fig. 6 Alpha
US-08-904-871-3

```

```

Query Match          65.6%; Score 672; DB 3; Length 600;
Best Local Similarity 60.5%; Pred. No. 2,9e-76;
Matches 127; Conservative 28; Mismatches 41; Indels 14; Gaps 1;

```

```

Qy 1 KLAVAISRLQSLPGDIGALCDTVEEDVQRLTGYDRVAVVYQFHEDDGEVVSSEIRSDL 60
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 202 KLAARAIISRLQSLPGDIGALCDTVEEDVQRLTGYDRVAVVYQFHEDDGEVVSSEIRSDL 261
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 61 EPLYGLHYPATDIPQARFLFKQNRVIMICDNCATPPVKVYQSEELKRPCLVNSTLRAPH 120
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 263 EPLYGLHYPATDIPQARFLFKQNRVIMICDNCATPPVKVYQSEELKRPCLVNSTLRAPH 321
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 121 GCHTOYMANMGVSASIALAIYVKGDS-----SKLMGLVVGHHCSPPRYVPF 166
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 322 SCHLOYMANMGVSASIALAIYVKGDS-----SKLMGLVVGHHCSPPRYVPF 381
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 167 PLRYACEFLMOAFGLQLOMELQASOLAER 196
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 382 PLRYACEFLMOAFGLQLOMELQASOLAER 411
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

RESULT 5
US-08-904-871-2
; Sequence 2, Application US/08904871
; Patent No. 6046014
; GENERAL INFORMATION:
; APPLICANT: Lagarias, John C
; APPLICANT: Murphy, John T
; TITLE OF INVENTION: PHYTOFLUORS AS FLUORESCENT LABELS
; FILE REFERENCE: 2500.134USO UC OTT Lagarias Patent
; CURRENT APPLICATION NUMBER: US/08/904,871
; CURRENT FILING DATE: 1997-08-01
; EARLIER APPLICATION NUMBER: 60/023,217
; EARLIER FILING DATE: 1996-08-02
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent In Ver. 2.0

```

```
; SEQ ID NO 2
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Sequence
US-08-904-871-2
```

```
Query Match          64.8%; Score 663.5; DB 3; Length 600;
Best Local Similarity 59.7%; Pred. No. 3.4e-75;
Matches 126; Conservative 24; Mismatches 46; Indels 15; Gaps 1;
```

```
QY 1 KLAVALAISRLQSLPGDICALCTVVEDVQRLTGYDRVMVYQFHEDDHGEVSEIRSDL 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 201 KLAVALAISRLQSLPGSMSEVLCNTVYKEVFDLTGYDRVMVYFHHDDHGEVSEITKPEL 260
QY 61 EPLYGLHYPATDIPQAAFLFKQNRVMIICDGNATPVKVYVSEELKRPCLVNSTLRAPH 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 261 EPLYGLHYPATDIPQAAFLFKQNRVMIICDGNATPVKVYVSEELKRPCLVNSTLRAPH 320
QY 121 GCHTOYMANMGSVASLALAIYVKGKDSKLMGLVYVGHHCSPRYVP 165
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 321 SCHLOYMENMNSIASLVAVVYVNEDEDEASEQPAQOQKKKKLWGLVCHHESPRYPV 380
```

```
QY 166 FPLRYACEFLMQAFGLQJOMELQJLASQJAEK 196
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 381 FPLRYACEFLMQAFVAVHVRREFELEKJLREK 411
```

```
RESULT 6
US-08-904-871-11
; Sequence 11, Application US/08904871
; Patent No. 6046014
; GENERAL INFORMATION:
; APPLICANT: Lagarias, John T
; TITLE OF INVENTION: PHYTOFLUORS AS FLUORESCENT LABELS
; FILE REFERENCE: 2500.134US0 UC OTT Lagarias Patent
; CURRENT APPLICATION NUMBER: US/08/904,871
; EARLIER FILING DATE: 1997-08-01
; EARLIER APPLICATION NUMBER: 60/023,217
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1129
; TYPE: PRT
; ORGANISM: Oat (Avena)
US-08-904-871-11
```

```
Query Match          64.8%; Score 663.5; DB 3; Length 1129;
Best Local Similarity 59.7%; Pred. No. 9.1e-75;
Matches 126; Conservative 24; Mismatches 46; Indels 15; Gaps 1;
```

```
QY 1 KLAVALAISRLQSLPGDICALCTVVEDVQRLTGYDRVMVYQFHEDDHGEVSEIRSDL 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 201 KLAVALAISRLQSLPGSMSEVLCNTVYKEVFDLTGYDRVMVYFHHDDHGEVSEITKPEL 260
QY 61 EPLYGLHYPATDIPQAAFLFKQNRVMIICDGNATPVKVYVSEELKRPCLVNSTLRAPH 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 261 EPLYGLHYPATDIPQAAFLFKQNRVMIICDGNATPVKVYVSEELKRPCLVNSTLRAPH 320
QY 121 GCHTOYMANMGSVASLALAIYVKGKDSKLMGLVYVGHHCSPRYVP 165
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 321 SCHLOYMENMNSIASLVAVVYVNEDEDEASEQPAQOQKKKKLWGLVCHHESPRYPV 380
QY 166 FPLRYACEFLMQAFGLQJOMELQJLASQJAEK 196
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 381 FPLRYACEFLMQAFVAVHVRREFELEKJLREK 411
```

```
RESULT 7
US-08-904-871-5
; Sequence 5, Application US/08904871
; Patent No. 6046014
; GENERAL INFORMATION:
; APPLICANT: Lagarias, John T
; TITLE OF INVENTION: PHYTOFLUORS AS FLUORESCENT LABELS
; FILE REFERENCE: 2500.134US0 UC OTT Lagarias Patent
; CURRENT APPLICATION NUMBER: US/08/904,871
; EARLIER FILING DATE: 1997-08-01
; EARLIER APPLICATION NUMBER: 60/023,217
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 528
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Sequence
US-08-904-871-5
```

```
Query Match          37.9%; Score 388; DB 3; Length 528;
Best Local Similarity 44.0%; Pred. No. 2.2e-40;
Matches 77; Conservative 37; Mismatches 53; Indels 8; Gaps 3;
```

```
QY 2 LAVRAISRLOSLPGDICALCTVVEDVQRLTGYDRVMVYQFHEDDHGEVSEIRSDLE 61
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 139 MANAALNRLRQ--QANLRPEFDVYVEVRMTGFDVMLRFDENHGVIAEDKRDME 196
QY 62 EPLYGLHYPATDIPQAAFLFKQNRVMIICDGNATPVKVYVSEELKRPCLVNSTLRAP 119
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 197 EPLYGLHYPATDIPQAAFLFKQNRVMIICDGNATPVKVYVSEELKRPCLVNSTLRAP 256
QY 120 HCHTOYMANMGSVASLALAIYVKGKDSKLMGLVYVGHHCSPRYVPPLRYACEF 174
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 257 YCHHLYLTKMVGASLFTSLIKDG---HLWGLIACHHQTPRVIPFELRKACEF 307
```

```
RESULT 8
US-08-904-871-6
; Sequence 6, Application US/08904871
; Patent No. 6046014
; GENERAL INFORMATION:
; APPLICANT: Lagarias, John T
; TITLE OF INVENTION: PHYTOFLUORS AS FLUORESCENT LABELS
; FILE REFERENCE: 2500.134US0 UC OTT Lagarias Patent
; CURRENT APPLICATION NUMBER: US/08/904,871
; EARLIER FILING DATE: 1997-08-01
; EARLIER APPLICATION NUMBER: 60/023,217
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 748
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Fig 10B seq
US-08-904-871-6
```

```
Query Match          37.9%; Score 388; DB 3; Length 748;
Best Local Similarity 44.0%; Pred. No. 3.8e-40;
Matches 77; Conservative 37; Mismatches 53; Indels 8; Gaps 3;
```

```
QY 2 LAVRAISRLOSLPGDICALCTVVEDVQRLTGYDRVMVYQFHEDDHGEVSEIRSDLE 61
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Db 139 MANAALNLRLQ--QANLDFDYVIVEEVRMTGDFRVMILYRFDENNGHGVIAEDKRDDME 196  
Oy 62 PYLGHTPATDIPQARLEFKONRYRMICDCN--ATPVKVOSEELKRLPLCVNSTLRAP 119  
Db 197 PYLGHTYVESDIPQARLFHINPIRVIPDYGVAVPLTPAVNPSTNRVADLTESILRSA 256  
Oy 120 HGHCHQYMANNGSVASLALAIYVKGKSSKLMGLVYGHHCSPRYVFPPLRYACEF 174  
Db 257 YHCHLTLYLKNMGVGSALTSLIKDG----HLMGLIACHHQTQPKVIPFELRKACEF 307

RESULT 9  
US-08-904-871-13  
Sequence 13, Application US/08904871  
Patent No. 6046014  
GENERAL INFORMATION:  
APPLICANT: Lagarias, John C  
APPLICANT: Murphy, John T  
TITLE OF INVENTION: PHYTOFLUORS AS FLUORESCENT LABELS  
FILE REFERENCE: 2500.134050 UC OTT Lagarias Patent  
CURRENT APPLICATION NUMBER: US/08/904.871  
CURRENT FILING DATE: 1997-08-01  
EARLIER APPLICATION NUMBER: 60/023.217  
EARLIER FILING DATE: 1996-08-02  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 13  
LENGTH: 748  
TYPE: PRT  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Description of Unknown Organism: Cyanobacteria  
US-08-904-871-13

Query Match 37.9%; Score 388; DB 3; Length 748;  
Best Local Similarity 44.0%; Pred. No. 3.8e-40;  
Matches 77; Conservative 37; Mismatches 53; Indels 8; Gaps 3;

Oy 2 LAVRISRLOSIPGGDICALCDYVEDYORLTGYRVVWYOPFHEDDGEVSEIRSDLE 61  
Db 139 MANAALNLRLQ--QANLDFDYVIVEEVRMTGDFRVMILYRFDENNGHGVIAEDKRDDME 196  
Oy 62 PYLGHTPATDIPQARLEFKONRYRMICDCN--ATPVKVOSEELKRLPLCVNSTLRAP 119  
Db 197 PYLGHTYVESDIPQARLFHINPIRVIPDYGVAVPLTPAVNPSTNRVADLTESILRSA 256  
Oy 120 HGHCHQYMANNGSVASLALAIYVKGKSSKLMGLVYGHHCSPRYVFPPLRYACEF 174  
Db 257 YHCHLTLYLKNMGVGSALTSLIKDG----HLMGLIACHHQTQPKVIPFELRKACEF 307

RESULT 10  
US-08-197-792-43  
Sequence 43, Application US/08197792  
Patent No. 5525488  
GENERAL INFORMATION:  
APPLICANT: Anthony J. Mason  
APPLICANT: Peter H. Seeburg  
TITLE OF INVENTION: Nucleic Acid Encoding the Alpha or Beta Chains of Inhibin and  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 Inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/197.792  
FILING DATE: 16-FEB-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/958414  
FILING DATE: 08-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/744207  
FILING DATE: 12-AUG-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/215466  
FILING DATE: 05-JUL-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 06/906729  
FILING DATE: 31-DEC-1986  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 06/827710  
FILING DATE: 07-FEB-1986  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 06/783910  
FILING DATE: 03-OCT-1985  
ATTORNEY/AGENT INFORMATION:  
NAME: Hasak, Janet E.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: 297P2D4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1896  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 43:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 353 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-197-792-43

Query Match 7.1%; Score 72.5; DB 1; Length 353;  
Best Local Similarity 34.4%; Pred. No. 1.1;  
Matches 31; Conservative 14; Mismatches 30; Indels 15; Gaps 7;

Oy 37 RMYVQFHEHDEGEVSEI--RRSDLEPYLGHTYPATDIPQARLEFKONRYRMICD--C 92  
Db 144 RYKVV--FOEGHGDWRNNVYKRVYDLK--RSGWHTPLPTAIOA---LFRGERRNLNDVQC 198  
Oy 93 NA-----TPVKVVOSEELKRLPLCVNSTL 116  
Db 199 DSCQELAVPVFVDPGESHREPVVQARL 228

RESULT 11  
US-08-459-850-43  
Sequence 43, Application US/08459850  
Patent No. 5663568  
GENERAL INFORMATION:  
APPLICANT: Anthony J. Mason  
APPLICANT: Peter H. Seeburg  
TITLE OF INVENTION: Nucleic Acid Encoding the Alpha or  
TITLE OF INVENTION: Beta Chains of Inhibin and Method for Synthesizing Polypept  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 Inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible

```

1 OPERATING SYSTEM: PC-DOS/MS-DOS
2 SOFTWARE: patin (Genentech)
3 CURRENT APPLICATION DATA:
4 APPLICATION NUMBER: US/08/459,850
5 FILING DATE: 02-JUN-1995
6 CLASSIFICATION: 435
7 PRIOR APPLICATION DATA:
8 APPLICATION NUMBER: 08/197792
9 FILING DATE: 17-FEB-1994
10 PRIOR APPLICATION DATA:
11 APPLICATION NUMBER: 07/958414
12 FILING DATE: 08-OCT-1992
13 PRIOR APPLICATION DATA:
14 APPLICATION NUMBER: 07/744207
15 FILING DATE: 12-AUG-1991
16 PRIOR APPLICATION DATA:
17 APPLICATION NUMBER: 07/215466
18 FILING DATE: 05-JUL-1988
19 PRIOR APPLICATION DATA:
20 APPLICATION NUMBER: 06/906729
21 FILING DATE: 31-DEC-1986
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: 06/827110
24 FILING DATE: 07-FEB-1986
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER: 06/783910
27 FILING DATE: 03-OCT-1985
28 ATTORNEY/AGENT INFORMATION:
29 NAME: Hasak, Janet E.
30 REGISTRATION NUMBER: 28,616
31 REFERENCE/DOCKET NUMBER: 297P2D5
32 TELECOMMUNICATION INFORMATION:
33 TELEPHONE: 415/225-1896
34 TELEFAX: 415/952-9881
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
96
```

```

Query March          . 7.1% Score 72.5; DB 1; Length 353;
Best Local Similarity 34.4%; Pred.No.1.1;
Matches 31; Conservative 14; Mismatches 30; Indels 15; Gaps 7;

QY      37 RVMVQFHHDEHGEVSEI-RRSIDLEPYLGILH-YPATDIPQAARLFKNRYRMICD--C 92
        ||| | | | : : : | | | : | : | | | | | : | |
Db       144 RVKVV-FOEGHGHDGMNNVEKRVDLK-RSGWHTTFPLTAIGA---LFRGERRLNDVQC 198

QY      93 NA-----TPVKVVOSEEKLRRPLCLVNSTL 116
        :: | | | | | : | : |
Db       199 DSCQELAVVPVFDPDGESHRRPFVVYQARL 228

RESULT 12
US-08-459-214-43
: Sequence 43, Application US/08459214
: Patent No. 5716810
GENERAL INFORMATION:
APPLICANT: Anthony J. Mason
APPLICANT: Peter H. Seeburg
TITLE OF INVENTION: Nucleic Acid Encoding the Alpha or
TITLE OF INVENTION: Beta Chains of Inhibin and Method for Synthesizing Polypeptide
TITLE OF INVENTION: using such Nucleic Acid
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA

```

```

1 ZIP: 944080
2 COMPUTER READABLE FORM:
3 MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
4 COMPUTER: IBM PC compatible
5 OPERATING SYSTEM: PC-DOS/MS-DOS
6 SOFTWARE: patin (Genentech)
7 CURRENT APPLICATION DATA:
8 APPLICATION NUMBER: US/08/459,214
9 FILING DATE: 02-JUN-1995
10 CLASSIFICATION: 435
11 PRIOR APPLICATION DATA:
12 APPLICATION NUMBER: 08/197792
13 FILING DATE: 17-FEB-1994
14 PRIOR APPLICATION DATA:
15 APPLICATION NUMBER: 07/958414
16 FILING DATE: 08-OCT-1992
17 PRIOR APPLICATION DATA:
18 APPLICATION NUMBER: 07/744207
19 FILING DATE: 12-AUG-1991
20 PRIOR APPLICATION DATA:
21 APPLICATION NUMBER: 07/215466
22 FILING DATE: 05-JUL-1988
23 PRIOR APPLICATION DATA:
24 APPLICATION NUMBER: 06/906729
25 FILING DATE: 31-DEC-1986
26 PRIOR APPLICATION DATA:
27 APPLICATION NUMBER: 06/827710
28 FILING DATE: 07-FEB-1986
29 PRIOR APPLICATION DATA:
30 APPLICATION NUMBER: 06/783910
31 FILING DATE: 03-OCT-1985
32 ATTORNEY/AGENT INFORMATION:
33 NAME: Hasak, Janet E.
34 REGISTRATION NUMBER: 28,616
35 REFERENCE/DOCKET NUMBER: 297P2D6
36 TELECOMMUNICATION INFORMATION:
37 TELEPHONE: 415/225-1896
38 TELEFAX: 415/952-9861
39
40 TELEX: 910/371-7168
41 INFORMATION FOR SEQ ID NO: 43:
42 SEQUENCE CHARACTERISTICS:
43 LENGTH: 353 amino acids
44 TYPE: amino acid
45 TOPOLOGY: linear
46
47 US-08-459-214-43

```

```

Query Match          7.1%: Score 72.5; DB 1; Length 353;
Best Local Similarity 34.4%: Pred. NO.1.1;
Matches 31; Conservative 14; Mismatches 30; Indels 15; Gaps
QY      37  RVMYVQFHHDDHGEVYSEI-RRSDELPYLGIA-YEATDIPQARFLFKONFRMYICD--C 92
          || || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      144  RVKVV-FEGQGHGDKWMNVVEKRVLDK-RSGWHTFPLTEALQA---LFEGERRLMIDVQC 198
QY      93  NA-----TPKVVYQSEELKRLPLCLVNSTL 116
          || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      199  DSCQELAVYVFPVDPGESHREPVVQAQL 228

RESULT 13
PCT-US93-05944-2
: Sequence 2, Application PC/TUS9305944
: GENERAL INFORMATION:
: APPLICANT: Lin et al., Hun-Chi
: TITLE OF INVENTION: Molecular cloning of the genes
: TITLE OF INVENTION: responsible for collagenase product
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Harris Brotman
: STREET: 401 B. St Ste 1700
: CITY: San Diego
: STATE: CA
:

```

COUNTRY: USA  
ZIP: 92101-4297  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/05944  
FILING DATE: 19930622  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brozman, Harris F.  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 699-3630  
TELEFAX: (619) 236-1048  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 936 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US93-05944-2

Query Match 7.0%; Score 72; DB 5; Length 936;  
Best Local Similarity 22.1%; Pred. No. 5.7;  
Matches 31; Conservative 19; Mismatches 62; Indels 28; Gaps 3;

QY 14 PGDIGALCDTVEDVQ---RLTGYDRVWYQFHHDDHGEVYSEIRRDLEPYLGLHYPA 70  
DB 711 PNDIKKANGPIVEGVYTKGDLNSDDADTFYFKEDGDVYIELPYGSSNFTLVLVKE 770  
QY 71 TDIDQARFLFKQNRVHMICDCAATPVKVOSEELKRPCLVNSTLRAPHCCHTOYMANM 130  
DB 771 GD-----DQNHIASGIDKNSKV-----GTFKATKGRHYFIYIKH 805  
QY 131 GSVASLALAIYVKGKSSKL 150  
DB 806 DSASNISYSLNKIGLGNEL 825

RESULT 14  
US-09-036-987A-6  
Sequence 6, Application US/09036987A  
Patent No. 6143526  
GENERAL INFORMATION:  
APPLICANT: Baltz, Richard H.  
APPLICANT: Broughton, Mary C.  
APPLICANT: Crawford, Kathryn P.  
APPLICANT: Madduri, Krishnamurthy  
APPLICANT: Merlo, Donald J.  
APPLICANT: Treadway, Patli J.  
APPLICANT: Turner, Jan R.  
APPLICANT: Waldron, Clive  
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dow Agrosciences LLC Patent Department  
STREET: 9330 Zionsville Road  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: USA  
ZIP: 46268  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/036,987A  
FILING DATE: 09-MAR-1998

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stuart, Donald R.  
REGISTRATION NUMBER: 28,479  
REFERENCE/DOCKET NUMBER: 50,608  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (317)337-4816  
TELEFAX: (317)337-4847  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5588 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-036-987A-6

Query Match 7.0%; Score 71.5; DB 4; Length 5588;  
Best Local Similarity 26.7%; Pred. No. 1e+02;  
Matches 44; Conservative 20; Mismatches 64; Indels 37; Gaps 8;

QY 4 VRAISRLOSLPGDIGALCDTVEDVQRLTGYDR---VMVYQFHHDDHGEVYSEIR-RSD 59  
DB 3445 IRRNRRASGTLEADGTLGVYREHAAVLTGSSADYGVERRARDLGFDSLGYELRNR 3504  
QY 60 LEPLYGLHYPATDI-----PQA-ARFLFKQNRVHMICDCAATPVKVOSEELKRPCLV 113  
DB 3505 LAGVLGVRLPATFVADYTPPALARFLHQE---LADRIATTPAPV----- 3546  
QY 114 STLRAPHCCHTOYMANMGSVASLALAIYVKGKSS--KLNLGVYV 156  
DB 3547 TTRAP-----VAEDDLVAIVGMCREFPGOVSSPEELMLVAG 3584

RESULT 15  
US-09-370-700-6  
Sequence 6, Application US/09370700  
Patent No. 6274350  
GENERAL INFORMATION:  
APPLICANT: Baltz, Richard H.  
APPLICANT: Broughton, Mary C.  
APPLICANT: Crawford, Kathryn P.  
APPLICANT: Madduri, Krishnamurthy  
APPLICANT: Treadway, Patli J.  
APPLICANT: Turner, Jan R.  
APPLICANT: Waldron, Clive  
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
FILE REFERENCE: 50489 DIV1  
CURRENT APPLICATION NUMBER: US/09/370,700  
CURRENT FILING DATE: 1999-08-09  
EARLIER APPLICATION NUMBER: US 09/36987  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: PatentIn ver. 2.0  
SEQ ID NO 6  
LENGTH: 5588  
TYPE: PRT  
ORGANISM: Saccharopolyspora spinosa  
US-09-370-700-6

Query Match 7.0%; Score 71.5; DB 4; Length 5588;  
Best Local Similarity 26.7%; Pred. No. 1e+02;  
Matches 44; Conservative 20; Mismatches 64; Indels 37; Gaps 8;

QY 4 VRAISRLOSLPGDIGALCDTVEDVQRLTGYDR---VMVYQFHHDDHGEVYSEIR-RSD 59  
DB 3445 IRRNRRASGTLEADGTLGVYREHAAVLTGSSADYGVERRARDLGFDSLGYELRNR 3504  
QY 60 LEPLYGLHYPATDI-----PQA-ARFLFKQNRVHMICDCAATPVKVOSEELKRPCLV 113  
DB 3505 LAGVLGVRLPATFVADYTPPALARFLHQE---LADRIATTPAPV----- 3546



OY 114 STLRAHGHCHQYMANMGSVASLALAIYVKGDSS--KLMLVYG 156  
:| | | : | | : | | : | | |  
Db 3547 TTTTAP-----VAEDDLVATVGCNRPFGVSSPEELMLVAG 3584

Search completed: June 7, 2002, 18:55:46  
Job time: 185 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2002, 18:56:32 ; Search time 37.51 Seconds  
(without alignments)  
502.093 million cell updates/sec

Title: US-09-272-809-9  
Perfect score: 1024  
Sequence: 1 KLAVRAISRQSLPGDGIGA.....QAFGLQLQMLQLASQLAEK 196

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues  
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1024	100.0	1112	2	S46313 phytochrome E - Ar
2	840.5	82.1	1131	2	T09701 phytochrome - Scot
3	836	81.6	1135	2	T16973 phytochrome B - cu
4	828	80.9	1132	2	T03668 phytochrome B - co
5	806	78.7	1039	2	T14802 phytochrome B - so
6	802.5	78.4	1172	1	EKMUB phytochrome B - Ar
7	794.5	77.6	1171	2	S14065 phytochrome B - ri
8	785	76.7	1142	2	S62714 phytochrome 1b - M
9	783	76.5	1164	2	B71429 phytochrome 1b - Ar
10	772	75.4	1156	2	T07756 phytochrome B - so
11	765.5	74.8	1136	2	T14842 phytochrome B - Norw
12	763.5	74.6	577	2	S58130 phytochrome - moss
13	763.5	74.6	1129	2	S28431 phytochrome B - po
14	747	72.9	1134	2	S31280 phytochrome B - Mart
15	746	72.9	368	2	T09496 phytochrome - Doug
16	740.5	72.3	1132	2	S37206 phytochrome - moss
17	734.5	71.7	1135	2	T14803 phytochrome C - so
18	723.5	70.7	1303	1	S27396 phytochrome BI - prot
19	717.5	70.1	211	1	S62720 phytochrome B1 - S
20	707	69.0	210	2	S62721 phytochrome B2 - S
21	706	68.9	1111	1	FKMUC phytochrome C - Ar
22	687	67.1	1465	2	T30891 PH3 protein - mai
23	672	65.6	1122	1	FKMUA phytochrome A - Ar
24	672	65.6	1122	2	D86229 phytochrome A [imp
25	671.5	65.6	1123	2	S20497 phytochrome A - po
26	669.5	65.4	1129	2	A29631 phytochrome A - oa
27	668.5	65.3	495	2	S00098 phytochrome 5 - oa
28	668	65.2	1129	2	S52631 phytochrome A - pa
29	666.5	65.1	1129	2	S00097 phytochrome 4 - oa

30	666	65.0	1124	1	FXPUZ phytochrome - zucc
31	665.5	65.0	1124	2	S06856 phytochrome - gard
32	665.5	65.0	1128	2	S03728 phytochrome (clone
33	662.5	64.7	1125	2	T09835 phytochrome A - po
34	660.5	64.5	1131	2	J00382 phytochrome A - ma
35	646	63.1	1131	2	T07137 phytochrome A - so
36	584.5	57.1	190	2	S46926 phytochrome - Mats
37	561	54.8	156	2	T14837 phytochrome 2 - NO
38	548	53.5	189	2	S46927 phytochrome - Nym
39	521	50.9	197	2	S46928 phytochrome - Char
40	513	50.1	156	2	T14838 phytochrome 3 - No
41	507.5	49.6	314	2	T07137 phytochrome - Norw
42	501	48.9	156	2	T14839 phytochrome 4 - No
43	489	47.8	156	2	T14836 phytochrome - Norw
44	474	46.3	105	2	T17026 phytochrome E - ga
45	455.5	44.5	115	2	T09337 phytochrome - Doug

## ALIGNMENTS

## RESULT 1

S46313  
phytochrome E - Arabidopsis thaliana  
N:Alternate names: protein F1505.100  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 27-Jan-1995 #sequence\_revision 27-Jan-1995 #text\_change 05-May-2000  
C:Accession: S46313; T14813; S41912  
R:Clack, T.; Mathews, S.; Sharrock, R.A.  
Plant Mol. Biol. 25, 413-427, 1994  
A:Title: The phytochrome apoprotein family in Arabidopsis is encoded by five genes: t  
A:Reference number: S46312; M01D:94325466  
A:Accession: S46313  
A:Molecule type: DNA  
A:Residues: 1-1112 <CLN>  
A:Cross-references: EMBL:Y76610; NID:9452815; PIDN:CA54075.1; PID:9452817  
R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.  
submitted to the Protein Sequence Database, August 1999  
A:Reference number: Z18120  
A:Molecule type: DNA  
A:Accession: T14813  
A:Residues: 1-1112 <BEV>  
A:Cross-references: EMBL:AL110123; GSPDB:GN00062; ATSP:F1505.100  
A:Experimental source: cultivar Columbia; BAC clone F1505  
C:Genetics:  
A:Gene: phyE; ATSP:F1505.100  
A:Map position: 4  
A:Introns: 665/1; 934/2; 1032/2  
C:Superfamily: phytochrome; phytochrome homology  
C:Keywords: chromoprotein; photoreceptor; phytochromobilin  
E:63-569/Domain: phytochrome homology <PHYT>  
F:322/Binding site: phytochromobilin (Cys) (covalent) #status predicted

Query Match 100.0%; Score 1024; DB 2; Length 1112;  
Best Local Similarity 100.0%; Pred. No. 1.8e-92;  
Matches 196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY	1	KLAVRAISRQSLPGDGIGA	CDTVVEDVQRLTGGRVVMVYQCHDDHGEVSEIRRS	DL 60
DB	201	KLAVRAISRQSLPGDGIGA	CDTVVEDVQRLTGGRVVMVYQCHDDHGEVSEIRRS	DL 260
OY	61	EPLYGLHYPATDIPQARELF	KONRVMICDCNATPVKVVQSELRKRLCLVNSTLRAPH	120
DB	261	EPLYGLHYPATDIPQARELF	KONRVMICDCNATPVKVVQSELRKRLCLVNSTLRAPH	320
OY	121	GCCTQYMANMGVAVSLALAI	IVKGRKSSKIMGLVGHHSKPRVPPPLRACFELMQAR	180
DB	321	GCCTQYMANMGVAVSLALAI	IVKGRKSSKIMGLVGHHSKPRVPPPLRACFELMQAR	380
OY	181	LQIMELQLASQLAEK	196	
DB	381	LQIMELQLASQLAEK	396	

## RESULT 2

phytochrome - Scotch pine

T09701

C:Species: Pinus sylvestris (Scotch pine)

C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 20-Jun-2000

C:Accession: T09701

R:Wiegmann-Eirund, C.M.; Kolukisaoglu, H.U.

Submitted to the EMBL Data Library, March 1996

A:Reference number: Z16826

A:Accession: T09701

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1131 &lt;WIE&gt;

A:Cross-references: EMBL:X96738; NID:g1237083

A:Experimental source: isolate PSA 5.1

C:Superfamily: phytochrome; phytochrome homology

C:Keywords: chromoprotein; photoreceptor; phytochromobilin

F:75-587/Domain: phytochrome homology &lt;PHY&gt;

F:332/Binding site: phytochromobilin (Cys) (covalent) #status predicted

## Query Match

Best Local Similarity 82.1%; Score 840.5; DB 2; Length 1131;

Matches 160; Conservative 15; Mismatches 21; Indels 9; Gaps 1;

OY 1 KLAVARAISRLQSLPGDGALCDTVVEDYQRLTGVDYRWVYGFHEDDGEVSEIRSDL 60

DB 211 KLAVARAISRLQSLPGDVGLLCDTVVENRELATGDRVWVYFHEDEHGEVVAESIRSDL 270

OY 61 EPYLGHYPATDIPQAAFLFKONRVKMDCAATPVKVOSEELKRPCLVNSTLRAPH 120

DB 271 EPYLGHYPATDIPQASRLFLKONRVKMDCAATPVKVOSEELKRPCLVNSTLRAPH 330

OY 121 GCHTOYMANMGVSASIALAIIVYKGD-----SSKLMGLVGHHCSPRYVPPFLRYAC 171

DB 331 GCHAOYMANMGVSIRSLAMVVIINGNDEGGSGSRMSKMLGLVGHHTSPRAVPPFLRYA 390

OY 172 CEFLMQAFGLQLOMELQIASQLAEK 196

DB 391 CEFLMQAFGLQLMELQIASQLAEK 415

## RESULT 3

phytochrome B - curled-leaved tobacco

T16973

C:Species: Nicotiana plumbaginifolia (curled-leaved tobacco)

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000

C:Accession: T16973

R:Hudson, M.E.; Robson, P.R.H.; Kraepiel, Y.; Caboche, M.; Smith, H.

Plant J. 12, 1091-2101, 1997

A:Title: Nicotiana plumbaginifolia hlg mutants have a mutation in a PHYB-type phytochrom

A:Reference number: Z18626; MUID:98079245

A:Accession: T16973

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1135 &lt;HUD&gt;

A:Cross-references: EMBL:Y14676; NID:g2370330; PIDN:CAA74992.1; PID:g2370331

A:Experimental source: cultivar PHND; leaf

C:Genetics:

A:Gene: PHYB

C:Superfamily: phytochrome; phytochrome homology

C:Keywords: chromoprotein; photoreceptor; phytochromobilin

F:82-592/Domain: phytochrome homology &lt;PHY&gt;

F:338/Binding site: phytochromobilin (Cys) (covalent) #status predicted

## Query Match

Best Local Similarity 81.6%; Score 836; DB 2; Length 1135;

Matches 157; Conservative 18; Mismatches 21; Indels 8; Gaps 1;

OY 1 KLAVARAISRLQSLPGDGALCDTVVEDYQRLTGDRVWVYGFHEDDGEVSEIRSDL 60

DB 217 KLAVARAISRLQSLPGDGVKILCDTVESVRELATGDRVWVYFHEDEHGEVVAESIRSDL 276

OY 61 EPYLGHYPATDIPQAAFLFKONRVKMDCAATPVKVOSEELKRPCLVNSTLRAPH 120

DB 277 EPYLGHYPATDIPQASRLFLKONRVKMDCAATPVKVOSEELKRPCLVNSTLRAPH 336

OY 121 GCHTOYMANMGVSASIALAIIVYKGD-----SSKLMGLVGHHCSPRYVPPFLRYAC 172

DB 337 GCHAOYMANMGVSIRSLAMVVIINGNDEAVGSRMSRLGLVGHHTSPRAVPPFLRYAC 396

OY 173 CEFLMQAFGLQLOMELQIASQLAEK 196

DB 397 CEFLMQAFGLQLMELQIASQLAEK 420

## RESULT 4

phytochrome B - common tobacco

T03668

C:Species: Nicotiana tabacum (common tobacco)

C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 21-Jan-2000

C:Accession: T03668; T03672

R:Kern, R.; Gasch, A.; Deak, M.; Kay, S.A.; Chua, N.H.

Plant Physiol. 102, 1363-1364, 1993

A:Title: PhyB of tobacco, a new member of the photoreceptor family.

A:Reference number: Z14996; MUID:94105358

A:Accession: T03668

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1132 &lt;KER&gt;

A:Cross-references: EMBL:L10114; NID:g295345; PIDN:AAA34092.1; PID:g295346

A:Experimental source: strain SR1; tissue-type etiolated seedling

R:Lopez-Juez, E.; Nagatani, A.; Tomizawa, K.; Deak, M.; Kern, R.; Kendrick, R.E.; Fur

Plant Cell 4, 241-251, 1992

A:Title: The cucumber long hypocotyl mutant lacks a light-stable PHYB-like phytochrom

A:Reference number: Z14997; MUID:92361250

A:Accession: T03672

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 457-506, 'S', 508-586, 'Q', 587-1132 &lt;LOP&gt;

A:Cross-references: EMBL:M65023; NID:g170286; PIDN:AAA34093.1; PID:g170287

C:Genetics:

A:Gene: PHYB

C:Superfamily: phytochrome; phytochrome homology

C:Keywords: chromoprotein; photoreceptor; phytochromobilin

F:80-589/Domain: phytochrome homology &lt;PHY&gt;

F:336/Binding site: phytochromobilin (Cys) (covalent) #status predicted

## Query Match

Best Local Similarity 80.9%; Score 828; DB 2; Length 1132;

Matches 156; Conservative 18; Mismatches 22; Indels 8; Gaps 1;

OY 1 KLAVARAISRLQSLPGDGALCDTVVEDYQRLTGDRVWVYGFHEDDGEVSEIRSDL 60

DB 215 KLAVARAISRLQSLPGDVGLLCDTVESVRELATGDRVWVYFHEDEHGEVVAESIRSDL 274

OY 61 EPYLGHYPATDIPQAAFLFKONRVKMDCAATPVKVOSEELKRPCLVNSTLRAPH 120

DB 275 EPYLGHYPATDIPQASRLFLKONRVKMDCAATPVKVOSEELKRPCLVNSTLRAPH 334

OY 121 GCHTOYMANMGVSASIALAIIVYKGD-----SSKLMGLVGHHCSPRYVPPFLRYAC 172

DB 335 GCHAOYMANMGVSIRSLAMVVIINGNDEAVGSRMSRLGLVGHHTSPRAVPPFLRYAC 394

OY 173 CEFLMQAFGLQLOMELQIASQLAEK 196

DB 395 CEFLMQAFGLQLMELQIASQLAEK 418

OY 5 KLAVARAISRLQSLPGDGALCDTVVEDYQRLTGDRVWVYGFHEDDGEVSEIRSDL 60

phytochrome B - sorghum (fragment)

C:/Species: Sorghum bicolor (sorghum)  
C:/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 05-May-2000  
C:/Accession: T14802  
R:/Childs, K.L.; Miller, F.R.; Cordunier-Pratt, M.M.; Pratt, L.H.; Morgan, P.W.; Mullet,  
submitted to the EMBL Data Library, April 1996  
C:/Description: The Sorghum bicolor photoperiod sensitivity gene, Ma3, encodes a phytochrome  
A:/Reference number: Z18185  
A:/Accession: T14802  
A:/Status: preliminary; translated from GB/EMBL/DDBT  
A:/Molecule type: DNA  
A:/Residues: 1-1039 <CHI>  
A:/Cross-references: EMBL:U56730; NID:g1800216; PID:g1800217  
A:/Experimental source: cultivar 58M  
C:/Genetics:  
A:/Gene: PHYB  
A:/Note: Intron positions not resolved (Incomplete sequence)  
C:/Superfamily: phytochrome; phytochrome homology  
C:/Keywords: chromoprotein; photoreceptor; phytochromobilin  
F:/233/Binding site: phytochromobilin (Cys) (covalent) #status predicted

Query Match 78.7%; Score 806; DB 2; Length 1039;  
Best Local Similarity 73.6%; Pred. No. 5.1e-71;  
Matches 153; Conservative 21; Mismatches 22; Indels 12; Gaps 2;

OY 1 KLARAISRLOSPLGCDIGALCPVVEDVOYLTCGYDRVMYQFHEDDHGEVSSELRSDL 60  
||||| : ||||| : ||||| : ||||| : ||||| :  
Db 112 KLARAISRLDALPGDGIKLLCDPVVEHRELGTGRVWYRFHDEHGEEVAESRRDNL 171  
  
OY 61 EPYGLIHPANDIPQARLEFKONVRMICNDGNPPVWVVOSEEKLRPCLVNSTLRAPH 120  
||||| : ||||| : ||||| : ||||| : ||||| :  
Db 172 EPLYGLIHAPADIPQASKEFLFRKNRVKMIADCHAPPAVIDPDGKSOPCLVGSTLRAPH 231  
  
OY 121 GCHTQYMNMGSVASLAIAIVK--GKD-----SKRMGLVSHGHSPRYPEPPL 168  
||| ||||| : ||| : | : : : : : : : : : : : : : : : : : : : : : : :  
Db 232 GCHAQGMANMGSIASLVMAVISSCGDDEQTGRGISAMKLMGLVVCNHNPSPIEPPL 291  
  
OY 169 RYACEFLMQAFGLQLOMELQLASQAER 196  
||||| : ||||| : ||||| : ||||| : ||||| :  
Db 292 RYACEFLMQAFGLQIMELQHLAQHSER 319

RESULT 6  
FCMDB  
phytochrome B - Arabidopsis thaliana  
C:/Species: Arabidopsis thaliana (mouse-ear cress)  
C:/Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 24-May-2001  
C:/Accession: B33473; JQ2141; F84568; S07718  
R:/Sharrock, R.A.; Quail, P.H.  
Genes Dev. 3, 1745-1757, 1989  
A:/Title: Novel phytochrome sequences in Arabidopsis thaliana: structure, evolution, and  
A:/Reference number: A33473; MUID:90108670  
A:/Accession: B33473  
A:/Molecule type: mRNA  
A:/Residues: 1-1172 <SHA>  
A:/Cross-references: EMBL:X17342; NID:g16422; PIDN:CA35222.1; PID:g16423  
R:/Reed, J.W.; Nagpal, P.; Poole, D.S.; Furuya, M.; Chory, J.  
Plant Cell 5, 147-157, 1993  
A:/Title: Mutations in the gene for the red/far-red light receptor phytochrome B alter ce  
A:/Reference number: JQ2141; MUID:93200802  
A:/Accession: JQ2141  
A:/Molecule type: DNA  
A:/Residues: 1-1172 <REE>  
A:/Cross-references: GB:L09262  
A:/Experimental source: ecotype Landsberg, mutant hy3  
R:/Lin, H.; Kaul, S.; Roundley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
Muss, D.; Nietman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,  
Nature 402, 761-768, 1999  
A:/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:/Reference number: A84420; MUID:20083487  
A:/Accession: F84568  
A:/Status: preliminary

Query Match	Best Local Similarity	Matches	Score	DB	Length
1	77.6%	71.8%	79.4.5	DB 2	1171
2	77.6%	71.8%	79.4.5	DB 2	1171
3	77.6%	71.8%	79.4.5	DB 2	1171
4	77.6%	71.8%	79.4.5	DB 2	1171
5	77.6%	71.8%	79.4.5	DB 2	1171
6	77.6%	71.8%	79.4.5	DB 2	1171
7	77.6%	71.8%	79.4.5	DB 2	1171
8	77.6%	71.8%	79.4.5	DB 2	1171
9	77.6%	71.8%	79.4.5	DB 2	1171
10	77.6%	71.8%	79.4.5	DB 2	1171
11	77.6%	71.8%	79.4.5	DB 2	1171
12	77.6%	71.8%	79.4.5	DB 2	1171
13	77.6%	71.8%	79.4.5	DB 2	1171
14	77.6%	71.8%	79.4.5	DB 2	1171
15	77.6%	71.8%	79.4.5	DB 2	1171
16	77.6%	71.8%	79.4.5	DB 2	1171
17	77.6%	71.8%	79.4.5	DB 2	1171
18	77.6%	71.8%	79.4.5	DB 2	1171
19	77.6%	71.8%	79.4.5	DB 2	1171
20	77.6%	71.8%	79.4.5	DB 2	1171
21	77.6%	71.8%	79.4.5	DB 2	1171
22	77.6%	71.8%	79.4.5	DB 2	1171
23	77.6%	71.8%	79.4.5	DB 2	1171
24	77.6%	71.8%	79.4.5	DB 2	1171
25	77.6%	71.8%	79.4.5	DB 2	1171
26	77.6%	71.8%	79.4.5	DB 2	1171
27	77.6%	71.8%	79.4.5	DB 2	1171
28	77.6%	71.8%	79.4.5	DB 2	1171
29	77.6%	71.8%	79.4.5	DB 2	1171
30	77.6%	71.8%	79.4.5	DB 2	1171
31	77.6%	71.8%	79.4.5	DB 2	1171
32	77.6%	71.8%	79.4.5	DB 2	1171
33	77.6%	71.8%	79.4.5	DB 2	1171
34	77.6%	71.8%	79.4.5	DB 2	1171
35	77.6%	71.8%	79.4.5	DB 2	1171
36	77.6%	71.8%	79.4.5	DB 2	1171
37	77.6%	71.8%	79.4.5	DB 2	1171
38	77.6%	71.8%	79.4.5	DB 2	1171
39	77.6%	71.8%	79.4.5	DB 2	1171
40	77.6%	71.8%	79.4.5	DB 2	1171
41	77.6%	71.8%	79.4.5	DB 2	1171
42	77.6%	71.8%	79.4.5	DB 2	1171
43	77.6%	71.8%	79.4.5	DB 2	1171
44	77.6%	71.8%	79.4.5	DB 2	1171
45	77.6%	71.8%	79.4.5	DB 2	1171
46	77.6%	71.8%	79.4.5	DB 2	1171
47	77.6%	71.8%	79.4.5	DB 2	1171
48	77.6%	71.8%	79.4.5	DB 2	1171
49	77.6%	71.8%	79.4.5	DB 2	1171
50	77.6%	71.8%	79.4.5	DB 2	1171
51	77.6%	71.8%	79.4.5	DB 2	1171
52	77.6%	71.8%	79.4.5	DB 2	1171
53	77.6%	71.8%	79.4.5	DB 2	1171
54	77.6%	71.8%	79.4.5	DB 2	1171
55	77.6%	71.8%	79.4.5	DB 2	1171
56	77.6%	71.8%	79.4.5	DB 2	1171
57	77.6%	71.8%	79.4.5	DB 2	1171
58	77.6%	71.8%	79.4.5	DB 2	1171
59	77.6%	71.8%	79.4.5	DB 2	1171
60	77.6%	71.8%	79.4.5	DB 2	1171
61	77.6%	71.8%	79.4.5	DB 2	1171
62					









GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2002, 18:57:39 ; Search time 60.43 Seconds  
(without alignments)  
561.096 Million cell updates/sec

Title: US-09-272-809-9  
Perfect score: 1024  
Sequence: 1 KLAIVRAISRQSLPGSDIGA.....QAFGLQLQMLQLASLAER 196

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SPREMBL\_19:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_nhc:\*  
9: sp\_organelle:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	894	87.3	1137	10	Q9M6P6 lycopersico
2	836.5	81.7	1130	10	Q24380 solanum tub
3	836.5	81.7	1131	10	Q9ZS62 lycopersico
4	836	81.6	1135	10	Q24117 nicotiana p
5	828	80.9	1121	10	Q9SWS6 lycopersico
6	823	80.4	1151	10	Q9FP03 populus trt
7	821	80.2	365	10	Q9FQW1 avertioha ca
8	817	79.8	365	10	Q9FQW4 goupia glab
9	817	79.8	365	10	Q9FQW0 avertioha ca
10	815	79.6	393	10	Q9XH84 sporobolus
11	813	79.4	245	10	Q9FQW3 avertioha ca
12	813	79.4	245	10	Q9FQW3 avertioha ca
13	813	79.4	374	10	Q9FQW3 avertioha ca
14	813	79.4	374	10	Q9FQW3 avertioha ca
15	813	79.4	388	10	Q9FQW3 avertioha ca
16	812.5	79.3	394	10	Q9XH94 pariana rad

17	812	79.3	245	10	Q9FQW5	Q9FQW5 goupia glab
18	811	79.2	245	10	Q9FQW5	Q9FQW5 salacia imp
19	811	79.2	245	10	Q9FQW5	Q9FQW5 salacia imp
20	811	79.2	388	10	Q9XH89	Q9XH89 phragmites
21	810	79.1	245	10	Q9FQW6	Q9FQW6 morionia gr
22	809.5	79.1	392	10	Q9XH84	Q9XH84 lithacne p
23	809	79.0	361	10	Q9FQW0	Q9FQW0 periperygi
24	809	79.0	393	10	Q9XH85	Q9XH85 hakonechloa
25	806	78.7	245	10	Q9FQW9	Q9FQW9 eucryphia b
26	806	78.7	382	10	Q9XH85	Q9XH85 sorghum hal
27	806	78.7	1146	10	Q9FQW2	Q9FQW2 populus tri
28	805	78.6	245	10	Q9FQW7	Q9FQW7 aleurites m
29	805	78.6	393	10	Q9XH81	Q9XH81 dantonionops
30	805	78.6	393	10	Q9XH85	Q9XH85 panticum cap
31	804	78.5	245	10	Q9FQW8	Q9FQW8 gymnosporia
32	804	78.5	245	10	Q9FQW4	Q9FQW4 paxistima c
33	804	78.5	245	10	Q9FQW3	Q9FQW3 paxistima c
34	804	78.5	245	10	Q9FQW0	Q9FQW0 paxistima m
35	803.5	78.5	394	10	Q9XH80	Q9XH80 anomochloa
36	803.5	78.5	394	10	Q9XH86	Q9XH86 olyra latif
37	803	78.4	245	10	Q9FQW9	Q9FQW9 gymnosporia
38	803	78.4	363	10	Q9FQW3	Q9FQW3 cuervea kap
39	803	78.4	365	10	Q9FQW1	Q9FQW1 loeseneriel
40	803	78.4	374	10	Q9FQW1	Q9FQW1 psammomya
41	802	78.3	245	10	Q9FQW5	Q9FQW5 putterlicki
42	802	78.3	245	10	Q9FQW2	Q9FQW2 avertioha ca
43	802	78.3	360	10	Q9FQW8	Q9FQW8 wimmeria ac
44	802	78.3	374	10	Q9FQW4	Q9FQW4 cuervea int
45	802	78.3	374	10	Q9FQW7	Q9FQW7 reissantia

## ALIGNMENTS

RESULT	ID	Q9M6P6	PRELIMINARY:	PRT:	1137 AA.
AC	Q9M6P6	01-OCT-2000 (TREMBLREL. 15, Created)			
DT	01-OCT-2000 (TREMBLREL. 15, Last sequence update)				
DT	01-DEC-2001 (TREMBLREL. 19, Last annotation update)				
DE	PHYTOCHROME E.				
GN	PHYE.				
OS	Lycopersicon esculentum (Tomato).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
OC	Asteridae; euasterids I; Solanales; Solanaceae; Solanum.				
OX	NCBI_TaxID=4081;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE-20188796; PubMed-10723737;				
RA	Alba R., Kelmenson P.M., Cordonnier-Pratt M.-M., Pratt L.H.;				
RT	"The phytochrome gene family in tomato and the rapid differential				
RT	evolution of this family in angiosperms.";				
RT	Mol. Biol. Evol. 17:362-373(2000).				
RL	EMBL, AF178571; AAP25812.1; -				
DR	InterPro: IPR003018; GAF.				
DR	InterPro: IPR003594; HATPase_c.				
DR	InterPro: IPR003661; His_kin.				
DR	InterPro: IPR004359; His_KIN_s1g.				
DR	InterPro: IPR000014; PAS				
DR	InterPro: IPR000700; PAS-assoc.C.				
DR	InterPro: IPR001294; Phytochrome.				
DR	Pfam: PF01590; GAF; 1.				
DR	Pfam: PF02518; HATPase_c; 1.				
DR	Pfam: PF00989; PAS; 2.				
DR	Pfam: PF00360; phytochrome; 1.				
DR	Pfam: PF00512; signal; 1.				
DR	PRINTS: PR01033; PHYTOCHROME.				
DR	SMART: SM00065; GAF; 1.				
DR	SMART: SM00387; HATPase_c; 1.				
DR	SMART: SM00388; HISKA; 1.				
DR	SMART: SM00091; PAS; 2.				







DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE PHYTOCHROME B (FRAGMENT).  
GN PHB.  
OS Goupia glabra.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids I; Celastraceae; Goupia.  
OX NCBI\_TaxID=39314;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Simmons M.P., Clevinger C.C., Savolainen V., Archer R.H., Mathews S.,  
RA Doyle J.J.;  
RT "Phylogeny of the Celastraceae Inferred from phytochrome B and  
RT morphology."  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF216086; AAC49043.1; -  
DR InterPro: IPR003018; GAF.  
DR InterPro: IPR001294; Phytochrome.  
DR Pfam: PF01590; GAF; 1.  
DR Pfam: PF00360; phytochrome; 1.  
DR PRINTS: PR01033; PHYTOCHROME.  
DR SMART: SM00065; GAF; 1.  
DR PROSITE: PS00245; PHYTOCHROME\_1; 1.  
DR PROSITE: PS50046; PHYTOCHROME\_2; 1.  
FT NON\_TER 1  
FT SEQUENCE 245 AA; 27109 MW; 55F2E41CC97B30B CRC64;

Query Match 79.8%; Score 817; DB 10; Length 245;  
Best Local Similarity 75.5%; Pred. No. 1.4e-76;  
Matches 154; Conservative 21; Mismatches 21; Indels 8; Gaps 1;  
QY 1 KLAVRAISRLOSIPGDIAGLCTVYEDVQRLTGYDRVWVYOFHEDHGEVSEIRSDL 60  
Db 12 KLAVRAISRLOSIPGDIAGLCTVYEDVQRLTGYDRVWVYOFHEDHGEVSEIRSDL 71  
QY 61 EPYGLHYPATDIPQAAFLFKQNRVMIICDCAIPVYVVOSEELKRPICLVNSTLRAPH 120  
Db 72 EPYGLHYPATDIPQAAFLFKQNRVMIICDCAIPVYVVOSEELKRPICLVNSTLRAPH 131  
QY 121 GCHQYMANMGSVASLAIATVYVKGD-----SSKIMGLVGHGCSPRVYFPLRYAC 172  
Db 132 GCHQYMANMGSVASLAIATVYVKGD-----SSKIMGLVGHGCSPRVYFPLRYAC 191  
QY 173 EFLMQAFGLQLOMELQJLASQLAEK 196  
Db 192 EFLMQAFGLQLOMELQJLASQLAEK 215

RESULT 9  
Q9FQW0 ID Q9FQW0 PRELIMINARY; PRT; 365 AA.  
AC Q9FQW0;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE PHYTOCHROME B (FRAGMENT).  
GN PHB.  
OS Averrhoa carambola (Star fruit).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids I; Oxalidaceae; Oxalidaceae; Averrhoa.  
OX NCBI\_TaxID=28974;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Simmons M.P., Clevinger C.C., Savolainen V., Archer R.H., Mathews S.,  
RA Doyle J.J.;  
RT "Phylogeny of the Celastraceae Inferred from phytochrome B and  
RT morphology."  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF216092; AAC49049.1; -  
DR InterPro: IPR000977; DNA\_Ligase.  
DR InterPro: IPR003018; GAF.  
DR InterPro: IPR001294; Phytochrome.  
DR Pfam: PF01590; GAF; 1.  
DR Pfam: PF00360; phytochrome; 1.  
DR PRINTS: PR01033; PHYTOCHROME.  
DR SMART: SM00065; GAF; 1.  
DR PROSITE: PS00245; DNA\_LIGASE\_A1; UNKNOWN\_1.  
DR PROSITE: PS50046; PHYTOCHROME\_2; 1.  
FT NON\_TER 1  
FT SEQUENCE 365 AA; 40359 MW; B464987BCE2E013C CRC64;

Query Match 79.8%; Score 817; DB 10; Length 365;  
Best Local Similarity 75.5%; Pred. No. 2.3e-76;  
Matches 154; Conservative 19; Mismatches 23; Indels 8; Gaps 1;  
QY 1 KLAVRAISRLOSIPGDIAGLCTVYEDVQRLTGYDRVWVYOFHEDHGEVSEIRSDL 60  
Db 12 KLAVRAISRLOSIPGDIAGLCTVYEDVQRLTGYDRVWVYOFHEDHGEVSEIRSDL 71  
QY 61 EPYGLHYPATDIPQAAFLFKQNRVMIICDCAIPVYVVOSEELKRPICLVNSTLRAPH 120  
Db 72 EPYGLHYPATDIPQAAFLFKQNRVMIICDCAIPVYVVOSEELKRPICLVNSTLRAPH 131  
QY 121 GCHQYMANMGSVASLAIATVYVKGD-----SSKIMGLVGHGCSPRVYFPLRYAC 172  
Db 132 GCHQYMANMGSVASLAIATVYVKGD-----SSKIMGLVGHGCSPRVYFPLRYAC 191  
QY 173 EFLMQAFGLQLOMELQJLASQLAEK 196  
Db 192 EFLMQAFGLQLOMELQJLASQLAEK 215

RESULT 10  
Q9XH84 ID Q9XH84 PRELIMINARY; PRT; 393 AA.  
AC Q9XH84;  
DT 01-NOV-1999 (Tremblrel. 12, Created)  
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE PHYTOCHROME B (FRAGMENT).  
GN Sporobolus giganteus.  
OS Sporobolus giganteus.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
OC Chloridaceae; Eragrostidae; Sporobolus.  
OX NCBI\_TaxID=96053;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Mathews S.Y., Tsai R.C., Kellogg E.A.;  
RT "Phylogenetic structure in the grass family (Poaceae): evidence from  
RT the nuclear gene phytochrome B."  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF137327; AAD41315.1; -  
DR InterPro: IPR000977; DNA\_Ligase.  
DR InterPro: IPR003018; GAF.  
DR InterPro: IPR001294; Phytochrome.  
DR Pfam: PF01590; GAF; 1.  
DR Pfam: PF00360; phytochrome; 1.  
DR PRINTS: PR01033; PHYTOCHROME.  
DR SMART: SM00065; GAF; 1.  
DR PROSITE: PS00697; DNA\_LIGASE\_A1; UNKNOWN\_1.  
DR PROSITE: PS00245; PHYTOCHROME\_1; 1.  
DR PROSITE: PS50046; PHYTOCHROME\_2; 1.  
FT NON\_TER 1  
FT SEQUENCE 393 AA; 43566 MW; 119CABFACFD4B13 CRC64;

Query Match 79.6%; Score 815; DB 10; Length 393;

Best Local Similarity 75.0%; Pred. No. 4.1e-76;  
Matches 156; Conservative 19; Mismatches 21; Indels 12; Gaps 2;

QY 1 KLAVALAISRLQSLPGDICALDPTVEYQRLTGYDRWVYQFHEDDGEVYSEIRSDL 60  
DB 20 KLAVALAISRLQSLPGDICALDPTVEYQRLTGYDRWVYQFHEDDGEVYSEIRSDL 79  
QY 61 EPIGLHYPATDIPQASRFLKONRVRMICDNATPVVVOSEELKRLPLCVNSTLRAPH 120  
DB 80 EPIGLHYPATDIPQASRFLKONRVRMICDNATPVVVOSEELKRLPLCVNSTLRAPH 139  
QY 121 GCHQYMANMGSVASLALAIYVKGKDS-----SKLMGLVGHCSPRVYPPPLR 168  
DB 140 GCHQYMANMGSVASLALAIYVKGKDS-----SKLMGLVGHCSPRVYPPPLR 199  
QY 169 RYACEFLMQAFGLQLOMELQLASQLAEK 196  
DB 200 RYACEFLMQAFGLQLOMELQLASQLAEK 227

RESULT 11  
Q9F0M3 ID Q9F0M3 PRELIMINARY; PRT; 245 AA.  
AC Q9F0M3  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE PHYTOCHROME B (FRAGMENT).  
GN PHYB.  
OS Afrostrax sp. 'Cheek 5007 K'.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids I; Huaceae; Alstrotyax.  
OX NCBI\_TaxID=85715;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Simmons M.P., Clevinger C.C., Savolainen V., Archer R.H., Mathews S.,  
RA Doyle J.J.;  
RT "Phylogeny of the Celastraceae inferred from phytochrome B and  
RT morphology."  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF216087; AAC49044.1; -  
DR InterPro: IPR003018; GAF.  
DR InterPro: IPR001294; Phytochrome.  
DR Pfam: PF01590; GAF. 1.  
DR Pfam: PF00360; phytochrome. 1.  
DR PRINTS: PR01033; PHYTOCHROME.  
DR SMART: SM00065; GAF. 1.  
DR PROSITE: PS00245; PHYTOCHROME\_1; 1.  
DR PROSITE: PS0046; PHYTOCHROME\_2; 1.  
FT NON\_TER 1 1  
FT NON\_TER 245 245  
SO SEQUENCE 245 AA; 27112 MW; C2A7688CA996FD70 CRC64;

Query Match 79.4%; Score 813; DB 10; Length 245;  
Best Local Similarity 76.5%; Pred. No. 3.7e-76;  
Matches 156; Conservative 16; Mismatches 24; Indels 8; Gaps 1;

QY 1 KLAVALAISRLQSLPGDICALDPTVEYQRLTGYDRWVYQFHEDDGEVYSEIRSDL 60  
DB 12 KLAVALAISRLQSLPGDICALDPTVEYQRLTGYDRWVYQFHEDDGEVYSEIRSDL 71  
QY 61 EPIGLHYPATDIPQASRFLKONRVRMICDNATPVVVOSEELKRLPLCVNSTLRAPH 120  
DB 72 EPIGLHYPATDIPQASRFLKONRVRMICDNATPVVVOSEELKRLPLCVNSTLRAPH 131  
QY 121 GCHQYMANMGSVASLALAIYVKGKDS-----SKLMGLVGHCSPRVYPPPLR 172  
DB 132 GCHQYMANMGSVASLALAIYVKGKDS-----SKLMGLVGHCSPRVYPPPLR 191  
QY 173 EFLMQAFGLQLOMELQLASQLAEK 196  
DB 173 EFLMQAFGLQLOMELQLASQLAEK 227

DB 192 EFLMQAFGLQLOMELQLASQLAEK 215

RESULT 12  
Q9F0M5 ID Q9F0M5 PRELIMINARY; PRT; 245 AA.  
AC Q9F0M5  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE PHYTOCHROME B (FRAGMENT).  
GN PHYB.  
OS Mortonia greggii.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids I; Celastraceae; Mortonia.  
OX NCBI\_TaxID=123436;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Simmons M.P., Clevinger C.C., Savolainen V., Archer R.H., Mathews S.,  
RA Doyle J.J.;  
RT "Phylogeny of the Celastraceae inferred from phytochrome B and  
RT morphology."  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF216137; AAC49094.1; -  
DR InterPro: IPR003018; GAF.  
DR InterPro: IPR001294; Phytochrome.  
DR Pfam: PF01590; GAF. 1.  
DR Pfam: PF00360; phytochrome. 1.  
DR PRINTS: PR01033; PHYTOCHROME.  
DR SMART: SM00065; GAF. 1.  
DR PROSITE: PS00245; PHYTOCHROME\_1; 1.  
DR PROSITE: PS0046; PHYTOCHROME\_2; 1.  
FT NON\_TER 1 1  
FT NON\_TER 245 245  
SO SEQUENCE 245 AA; 26956 MW; E2AED28C0E76DCDF CRC64;

Query Match 79.4%; Score 813; DB 10; Length 245;  
Best Local Similarity 74.5%; Pred. No. 3.7e-76;  
Matches 152; Conservative 23; Mismatches 21; Indels 8; Gaps 1;

QY 1 KLAVALAISRLQSLPGDICALDPTVEYQRLTGYDRWVYQFHEDDGEVYSEIRSDL 60  
DB 12 KLAVALAISRLQSLPGDICALDPTVEYQRLTGYDRWVYQFHEDDGEVYSEIRSDL 71  
QY 61 EPIGLHYPATDIPQASRFLKONRVRMICDNATPVVVOSEELKRLPLCVNSTLRAPH 120  
DB 72 EPIGLHYPATDIPQASRFLKONRVRMICDNATPVVVOSEELKRLPLCVNSTLRAPH 131  
QY 121 GCHQYMANMGSVASLALAIYVKGKDS-----SKLMGLVGHCSPRVYPPPLR 172  
DB 132 GCHQYMANMGSVASLALAIYVKGKDS-----SKLMGLVGHCSPRVYPPPLR 191  
QY 173 EFLMQAFGLQLOMELQLASQLAEK 196  
DB 192 EFLMQAFGLQLOMELQLASQLAEK 215  
RESULT 13  
Q9F0Q9 ID Q9F0Q9 PRELIMINARY; PRT; 374 AA.  
AC Q9F0Q9  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE PHYTOCHROME B (FRAGMENT).  
GN PHYB.  
OS Perrotetia ovata.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OX eurosids I; Celastraceae; Perrotetia.  
OX NCBI\_TaxID=123447;



Mon Jun 10 09:44:16 2002

us-09-272-809-9.rspt

Page 8

```

Oy 121 GCHQYNNMMSVSLAIAIV-----KCKDSS--KIMELVYHHCSPPYVEPL 168
    ||| ||||| ||||| ||| :||| ||| ||||| ||| ||| ||| ||| ||| |||
Db 135 GCHQYNNMMSVSLAIAVAVIIITIGCDDEQYARSRISISAKMKIMLVYCHHTSPRFITPL 194
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 169 RYACEFLMQAFGLQLOMELQIAQLAELK 196
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 195 RYACEFLMQAFGLQINNELQIAHLDSEK 222
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Search completed: June 7, 2002, 18:57:40  
Job time: 239 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2002, 18:58:02 ; Search time 19.14 Seconds  
(without alignments)  
396.501 Million cell updates/sec

Title: US-09-272-809-9  
Sequence: 1 KTAVRAISRQLSGDIGH.....QAFGLDQLMEFLASOLAEX 196

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues  
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1024	100.0	1112	1	P42498 arabidopsis
2	914	89.3	1115	1	P55004 parhbitis n
3	840.5	82.1	1131	1	P41046 pinus sylve
4	828	80.9	1132	1	P28130 nicotiana t
5	806	78.7	1178	1	P39527 scorghum bic
6	802.5	78.4	1172	1	P14713 arabidopsis
7	794.5	77.6	1171	1	P25764 oryza sativ
8	783	76.5	1164	1	P42497 arabidopsis
9	772	75.4	1156	1	P42499 glycine max
10	765.5	74.8	1136	1	Q40762 picea abies
11	763.5	74.6	1121	1	Q39557 ceratodon p
12	753.5	74.6	1129	1	P34094 solanum tub
13	758.5	74.1	1118	1	P42496 adiantum ca
14	747	72.9	1134	1	O01505 selaginella
15	740.5	72.3	1132	1	P38505 physcomitre
16	734.5	71.7	1135	1	P39528 scorghum bic
17	732.5	70.7	1137	1	Q95W19 oryza sativ
18	723.5	70.7	1307	1	P25848 ceratodon p
19	722.5	70.6	1124	1	P33529 mengoetia s
20	706	68.9	1111	1	P14714 arabidopsis
21	672	65.6	1122	1	P14712 arabidopsis
22	671.5	65.6	1123	1	P30733 solanum tub
23	671.5	65.6	1124	1	P33530 solanum tub
24	668.5	65.3	494	1	P06595 avena sativ
25	668.5	65.3	1131	1	P33526 scorghum bic
26	668	65.2	1129	1	P06594 avena sativ
27	666.5	65.1	1128	1	P06592 cucurbita p
28	666	65.0	1124	1	P39673 latihyus sa
29	665.5	65.0	1124	1	P15001 pisum sativ
30	665.5	65.0	1124	1	P10931 oryza sativ
31	663.5	64.8	1128	1	P06593 oryza sativ
32	663.5	64.8	1128	1	P06593 oryza sativ
33	662.5	64.7	1125	1	O49934 populus tre

34	660.5	64.5	1131	1	PHYA_MAIZE	P19862 zea mays (m
35	646	63.1	1131	1	PHYA_SOYBN	P42500 glycine max
36	388	37.9	741	1	PHY1_SYNY3	Q55168 synecocyst
37	379	37.0	765	1	PHYA_ANASP	O91cc2 anabaena sp
38	366.5	35.8	728	1	BPHY_PSEAE	O9hrc3 pseudomonas
39	356	34.8	751	1	PHYB_AMASP	O9r6x3 anabaena sp
40	315	30.8	755	1	BPHY_DEIRA	O9r2a4 deinococcus
41	190	18.6	1276	1	PHY2_SYNY3	O55434 synecocyst
42	85.5	8.3	633	1	SIR_SYNY3	P72854 synecocyst
43	79	7.7	335	1	ETPA_RAT	P13803 rattus norv
44	79	7.7	989	1	RPOC_LEUME	P44892 leucocostoc
45	76	7.4	192	1	CCA_AERHY	P45749 aeromonas h

## ALIGNMENTS

RESULT 1  
ID PHYE\_ARATH STANDARD: PRT; 1112 AA.  
AC P42498;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Phytochrome E.  
GN PHYE OR AT4G18130 OR F15J5.100.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. LANDSBERG ERRECTA;  
RX MEDLINE=94325466; PubMed=8049367;  
RA "The phytochrome apoprotein family in Arabidopsis is encoded by five  
RT genes: the sequences and expression of PHYD and PHYE";  
RL Plant Mol. Biol. 25:413-427(1994).  
[2]  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RX MEDLINE=20083488; PubMed=10617198;  
RA Meyer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,  
RA Pohl T., Duesterhoeft A., Stiekema W., Entlan K.-D., Terryn N.,  
RA Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M.,  
RA Weichselgartner M., de Simone V., Obermayer B., Meche R., Mueller M.,  
RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidheini T.,  
RA Reichert B., Portetelle D., Perez-Alonso M., Bouly M., Bancroft I.,  
RA Vos P., Hohnsels J., Zimmermann W., Wedler H., Ridley P.,  
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,  
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,  
RA Breken M., Weltjens I., Voet M., Bastiens I., Aert R., Deltor E.,  
RA Weltzengerger T., Boche G., Ransperger U., Hilbert H., Braun M.,  
RA Holzner E., Brandt A., Peters S., Van Slaveren M., Ditsch W.,  
RA Woolman P., Klein Lankhorst R., Rose M., Haut J., Koetter P.,  
RA Bernreiser S., Hempel S., Feldpausch M., Lamberth S., Van den Deele H.,  
RA De Keyser A., Buysaert C., Gielen J., Villarroel R., De Clercq R.,  
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,  
RA Clark L., Doggett J., Hall S., Key M., Lennard N., McClay K., Mayes R.,  
RA Petlett A., Rajandream M.A., Lyne M., Benes V., Reichtman S.,  
RA Borikova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,  
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fartmann B., Grandprath K., Dauner D., Herzl A.,  
RA Neumann S., Argitlou A., Vitale D., Liguori R., Pivavandi E.,  
RA Massenet O., Ougley F., Clabaud G., Mendel A., Felber R.,  
RA Schnabl S., Hiller R., Schmidt W., Lechery A., Aubourg S.,  
RA Chedof F., Cooke R., Berger C., Montfort A., Casacuberta S.,  
RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,  
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,  
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Belke C.,  
RA Feilman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,  
RA Zaccaria P., Beyer M., Wilson R.K., de la Bastide M., Habermann K.,

RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,  
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Thaiden J.,  
 RA Stenking T., Rajicki J., Graves T., Harmon G., Edwards J.,  
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,  
 RA Mix P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,  
 RA Kramer J., Fulton L., Mardis E., Dance M., Pepin K., Hillier L.,  
 RA Nelson J., Speeth J., Ryan E., Andrews S., Geisel C., Layman D.,  
 RA Du H., All J., Berghoff A., Jones K., Drone K., Cotton M., Joshua C.,  
 RA Antonoulu B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,  
 RA Ma P., Zhong J., Preston R., Vill D., Shekhar M., Matero A., Shah R.,  
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,  
 RA Granet S., Shohdy N., Hasegawa A., Hamed A., Lodhi M., Johnson A.,  
 RA Chen E., Marra M., Martensen R., McComdie W.R.,  
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis  
 thaliana";  
 RL Nature 402:769-777(1999).  
 CC -1- FUNCTION: REGULATOR PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT  
 CC ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS  
 CC MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT  
 CC ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCOVERSION OF PR IN  
 CC PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS  
 CC RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE  
 CC RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR  
 CC GENES, INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-  
 CC BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,  
 CC PROTOCHLOROPHYLLIDE REDUCTASE, RNA, ETC. IT ALSO CONTROLS  
 CC THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- PTM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.  
 CC -1- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.  
 CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X76610; CAA54075.1; -;  
 DR EMBL: AL101023; CAB53654.1; -;  
 DR EMBL: AL161548; CAB78815.1; -;  
 DR Mendel: 7190; ARACH.PHYE;1.  
 DR InterPro: IPR003018; GAF.  
 DR InterPro: IPR003594; HATPase\_C.  
 DR InterPro: IPR004359; HIS\_KIN\_sig.  
 DR InterPro: IPR003661; HIS\_KINa.  
 DR InterPro: IPR000014; PAS.  
 DR InterPro: IPR001294; Phytochrome.  
 DR Pfam: PF01590; GAF; 1.  
 DR Pfam: PF02518; HATPase\_C; 1.  
 DR Pfam: PF00989; PAS; 2.  
 DR Pfam: PF00360; phytochrome; 1.  
 DR Pfam: PF00512; signal; 1.  
 DR PRINTS: PR01033; PHYTOCHROME.  
 DR SMART: SM00065; GAF; 1.  
 DR SMART: SM00387; HATPase\_C; 1.  
 DR SMART: SM00388; HSKA; 1.  
 DR SMART: SM00091; PAS; 2.  
 DR PROSITE: PS50109; HIS\_KIN; 1.  
 DR PROSITE: PS50112; PAS; 2.  
 DR PROSITE: PS00245; PHYTOCHROME\_1; 1.  
 DR PROSITE: PS50046; PHYTOCHROME\_2; 1.  
 KW Transcription regulation; Photoreceptor; Phytochrome; Chromophore;  
 KW Repeat: Multigene family.  
 FT DOMAIN 595 666 PAS 1.  
 FT DOMAIN 732 803 PAS 2.  
 FT DOMAIN 877 1096 HISTIDINE KINASE.  
 FT BINDING 322 322 CHROMOPHORE (BY SIMILARITY).  
 SQ SEQUENCE 1112 AA; 122587 MW; 4C235B83F6D9DA28 CRC64;

Query Match 100.0%; Score 1024; DB 1; Length 1112;  
 Best Local Similarity 100.0%; Pred. No. 6.6e-96;  
 Matches 196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KLAIVRAISLQSLPGDIGALCDTVYEDVQRLTGDRVYVQFHHDDHGEVSEIRSDL 60  
 DB 201 KLAIVRAISLQSLPGDIGALCDTVYEDVQRLTGDRVYVQFHHDDHGEVSEIRSDL 260  
 QY 61 EYLGLHVPATIPQARFLFKONVRMTCDNATPVKVYVQSEELKRPCLVNSTLRAP 120  
 DB 261 EYLGLHVPATIPQARFLFKONVRMTCDNATPVKVYVQSEELKRPCLVNSTLRAP 320  
 QY 121 GCHTYMANMGSVASIALAIYVKGRDSSKLMGLVYGHHCSPRYVPPLRYACEFLMQAF 180  
 DB 321 GCHTYMANMGSVASIALAIYVKGRDSSKLMGLVYGHHCSPRYVPPLRYACEFLMQAF 380  
 QY 181 LQLOMELQASLAEK 196  
 DB 381 LQLOMELQASLAEK 396  
 RESULT 2  
 ID PHYE\_PHANI STANDARD; PRT; 1115 AA.  
 AC P55004;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DE 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Phytochrome E.  
 GN PHYE.  
 OS Pharbitis nil (Viollet) (Japanese morning glory).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; easterids I; Solanales; Convolvulaceae; Ipomoea.  
 OX NCBI\_TaxId=35883;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Seedling cotyledon;  
 RA Zheng C.C., O'Neill S.D.;  
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: REGULATOR PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT  
 CC ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS  
 CC MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT  
 CC ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCOVERSION OF PR IN  
 CC PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS  
 CC RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE  
 CC RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR  
 CC GENES, INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-  
 CC BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,  
 CC PROTOCHLOROPHYLLIDE REDUCTASE, RNA, ETC. IT ALSO CONTROLS  
 CC THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- PTM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.  
 CC -1- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.  
 CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U39787; AAA84970.1; -;  
 DR InterPro: IPR003018; GAF.  
 DR InterPro: IPR003594; HATPase\_C.  
 DR InterPro: IPR004359; HIS\_KIN\_sig.  
 DR InterPro: IPR003661; HIS\_KINa.  
 DR InterPro: IPR000700; PAS-ASSOC\_C.

DR InterPro: IPR000014; PAS.  
DR InterPro: IPR001294; Phytochrome.  
DR Pfam: PF01590; GAF; 1.  
DR Pfam: PF02518; HATPase\_C; 1.  
DR Pfam: PF00989; PAS; 2.  
DR Pfam: PF00360; phytochrome; 1.  
DR Pfam: PF00512; signal; 1.  
DR PRINTS: PRO1033; PHYTOCHROME.  
DR SMART: SM00065; GAF; 1.  
DR SMART: SM00387; HATPase\_C; 1.  
DR SMART: SM00091; PAS; 2.  
DR SMART: SM00091; PAS; 2.  
DR PROSITE: PS50109; HIS\_KIN; 1.  
DR PROSITE: PS50113; PAC; 1.  
DR PROSITE: PS50112; PAS; 2.  
DR PROSITE: PS00245; PHYTOCHROME\_1; 1.  
DR PROSITE: PS50046; PHYTOCHROME\_2; 1.  
KW Transcription regulation; Photoreceptor; Phytochrome; Chromophore;  
KW Repeat.  
FT DOMAIN 598 669 PAS 1.  
FT DOMAIN 672 728 PAC.  
FT DOMAIN 732 803 PAS 2.  
FT DOMAIN 880 1100 HISTIDINE KINASE.  
FT BINDING 318 318 CHROMOPHORE (BY SIMILARITY).  
SQ SEQUENCE 1115 AA: 124328 MW: 081A4154E147800 CRC64;

Query Match 89.3%; Score 914; DB 1; Length 1115;  
Best Local Similarity 87.2%; Pred. No. 9.7e-85;  
Matches 171; Conservative 12; Mismatches 13; Indels 0; Gaps 0;

QY 1 KLAVAIRISLQSLPGDGLCDTVYEDVQRLTGDRVWVYGFHEDHGEVSEIRSL 60  
DB 197 KLAVAIRISLQSLPGDGLCDTVYEDVQRLTGDRVWVYGFHEDHGEVSEIRSL 256  
QY 61 EPYGLHYATDIPQARFLFKONRYRMICDCAATPVKVOSEELKRPCLVNSTLRAP 120  
DB 257 EPYGLHYATDIPQARFLFKONRYRMICDCAATPVKVOSEELKRPCLVNSTLRAP 316  
QY 121 GCHTOYMANMGSVASLAIIVKGRDSSKLMGLVGVHHCSPRYVPPPLRYA 180  
DB 317 GCHTOYMANMGSVASLAIIVKGRDSSKLMGLVGVHHCSPRYVPPPLRYA 376  
QY 181 LQLMQELQSLAQLEK 196  
DB 377 LQLMQELQSLAQLEK 392

RESULT 3  
PHY\_PINSY STANDARD; PRT; 1131 AA.  
ID PHY\_PINSY  
AC 041046;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Phytochrome.  
OS Pinus sylvestris (Scots pine).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.  
OX NCBI\_TaxID=3349;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PS 5.1;  
RA Wiegman-Eirund C.M., Kolukisaoglu H.U.;  
RL Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.  
CC -1- FUNCTION: REGULATORY PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT  
CC ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS  
CC MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT  
CC ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCOVERSION OF PR IN  
CC PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS  
CC RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE  
CC RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR  
CC GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-

CC BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,  
CC PROTOCHLOROPHYLLIDE REDUCTASE, RNA, ETC. IT ALSO CONTROLS  
CC THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION.  
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
CC -1- PFM: CONTAINS ONE COVALENTLY LINKED TETRAPHYRROLE CHROMOPHORE.  
CC -1- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.  
CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ANT-SIM) DIMERIZATION DOMAINS.  
CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sb.ch/announce/>  
CC or send an email to [license@isb-sb.ch](mailto:license@isb-sb.ch)).  
CC -----  
DR EMBL: X96738; CAA65510.1; -.  
DR InterPro: IPR003018; GAF.  
DR InterPro: IPR003594; HATPase\_C.  
DR InterPro: IPR004359; HIS\_KIN\_sig.  
DR InterPro: IPR003661; HIS\_KIN.  
DR InterPro: IPR001610; PAC.  
DR InterPro: IPR000014; PAS.  
DR InterPro: IPR001294; Phytochrome.  
DR Pfam: PF01590; GAF; 1.  
DR Pfam: PF02518; HATPase\_C; 1.  
DR Pfam: PF00989; PAS; 2.  
DR Pfam: PF00360; phytochrome; 1.  
DR Pfam: PF00512; signal; 1.  
DR PRINTS: PRO1033; PHYTOCHROME.  
DR SMART: SM00065; GAF; 1.  
DR SMART: SM00387; HATPase\_C; 1.  
DR SMART: SM00086; HIS\_KIN; 1.  
DR SMART: SM00091; PAS; 2.  
DR PROSITE: PS50109; HIS\_KIN; 1.  
DR PROSITE: PS50112; PAS; 2.  
DR PROSITE: PS00245; PHYTOCHROME\_1; 1.  
DR PROSITE: PS50046; PHYTOCHROME\_2; 1.  
KW Transcription regulation; Photoreceptor; Phytochrome; Chromophore;  
KW Repeat.  
FT DOMAIN 621 692 PAS 1.  
FT DOMAIN 755 826 PAS 2.  
FT DOMAIN 903 1123 HISTIDINE KINASE.  
FT BINDING 332 332 CHROMOPHORE (BY SIMILARITY).  
SQ SEQUENCE 1131 AA: 126254 MW: D63A2008FA9862EB CRC64;

Query Match 82.1%; Score 840.5; DB 1; Length 1131;  
Best Local Similarity 78.0%; Pred. No. 2.9e-77;  
Matches 160; Conservative 15; Mismatches 21; Indels 9; Gaps 1;

QY 1 KLAVAIRISLQSLPGDGLCDTVYEDVQRLTGDRVWVYGFHEDHGEVSEIRSL 60  
DB 211 KLAVAIRISLQSLPGDGLCDTVYEDVQRLTGDRVWVYGFHEDHGEVSEIRSL 270  
QY 61 EPYGLHYATDIPQARFLFKONRYRMICDCAATPVKVOSEELKRPCLVNSTLRAP 120  
DB 271 EPYGLHYATDIPQARFLFKONRYRMICDCAATPVKVOSEELKRPCLVNSTLRAP 330  
QY 121 GCHTOYMANMGSVASLAIIVKGRDSSKLMGLVGVHHCSPRYVPPPLRYA 171  
DB 331 GCHTOYMANMGSVASLAIIVKGRDSSKLMGLVGVHHCSPRYVPPPLRYA 390  
QY 172 CEFMQARGLQLMQELQSLAQLEK 196  
DB 391 CEFMQARGLQLMQELQSLAQLEK 415

RESULT 4  
PHYB\_TOBAC

ID PHYB\_TOBAC STANDARD: PRT, 1132 AA.  
AC P29130;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Phytochrome B.  
GN PHYB.  
OS Nicotiana tabacum (Common tobacco).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; eustersids I; Solanales; Solanaceae; Nicotiana.  
OX NCBI\_TaxID=4097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94105358; PubMed=8278560;  
RA Kern R., Gasch A., Deak M., Kay S.A., Chua N.H.;  
RT "Phyb of tobacco, a new member of the phytochrome family.";  
RL Plant Physiol. 102:1363-1364(1993).  
RN [2]  
RP SEQUENCE OF 457-1132 FROM N.A.  
RX MEDLINE=92361250; PubMed=1498594;  
RA Lopez-Juez E., Nagatani A., Tomizawa K.I., Deak M., Kern R.,  
RA Kendrick R.E., Furuya M.;  
RT "The cucumber long hypocotyl mutant lacks a light-stable PHYB-like  
phytochrome.";  
RL Plant Cell 4:241-251(1992).  
CC -1- FUNCTION: REGULATORY PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT  
ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS  
MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT  
ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCONVERSION OF PR IN  
PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS  
RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE  
RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR  
GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-  
BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,  
PHOTOCOLOROPHYLLIDE REDUCTASE, RNA, ETC. IT ALSO CONTROLS  
THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION.  
CC -1- SUBUNIT: HOMODIMER.  
CC -1- PFM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.  
CC -1- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.  
CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: L10114; AAA34092.1; -;  
DR EMBL: M65023; AAA34093.1; -;  
DR Mendel: 1321; NICta:PHYB.1.  
DR InterPro: IPR003018; GAF.  
DR InterPro: IPR003594; HATPase\_C.  
DR InterPro: IPR004359; HIS\_KIN\_sig.  
DR InterPro: IPR003661; HIS\_KINa.  
DR InterPro: IPR001610; PAC.  
DR InterPro: IPR000014; PAS.  
DR InterPro: IPR001294; Phytochrome.  
DR Pfam: PF01590; GAF; 1.  
DR Pfam: PF02518; HATPase\_C; 1.  
DR Pfam: PF00989; PAS; 2.  
DR Pfam: PF00360; phytochrome; 1.  
DR PRINTS: PF00512; signal; 1.  
DR PRINTS: PR01033; PHYTOCHROME.  
DR SMART: SM00065; GAF; 1.  
DR SMART: SM00387; HATPase\_C; 1.  
DR SMART: SM00388; HIS\_KA; 1.  
DR SMART: SM00086; PAC; 1.  
DR SMART: SM00091; PAS; 2.

DR PROSITE: PS50109; HIS\_KIN; 1.  
DR PROSITE: PS50112; PAS; 2.  
DR PROSITE: PS00245; PHYTOCHROME\_1; 1.  
DR PROSITE: PS50046; PHYTOCHROME\_2; 1.  
KW Transcription regulation; Photoreceptor; Phytochrome; Chromophore;  
KW Repeat; Multigene family.  
FT DOMAIN 623 694 PAS 1.  
FT DOMAIN 757 828 PAS 2.  
FT DOMAIN 905 1125 HISTIDINE KINASE.  
FT BINDING 336 336 CHROMOPHORE (BY SIMILARITY).  
FT CONFLICT 507 507 L -> S (IN REF. 2).  
FT CONFLICT 586 586 L -> LQ (IN REF. 2).  
SQ SEQUENCE 1132 AA; 125809 MW; 457F09C024C0F608 CRC64;  
Query Match 80.9%; Score 828; DB 1; Length 1132;  
Best local similarity 76.5%; Pred. No. 5,4e-76;  
Matches 156; Conservative 18; Mismatches 22; Indels 8; Gaps 1;  
QY 1 KLAVERAISRLQSLPGDIGALCDTVEDVQRLTGYDRVWVYQFHEDDGEVVSFIRSDL 60  
DB 215 KLAVERAISRLQSLPGDVKVLCDFVESVRELTYGDRVWVYKFHDEHGEVVAESKIPDL 274  
QY 61 EYILGHLPATDIPQARPLFKQNRVMTCDGNATPVKYVQSEELKRLPLCLYNSTLRAPH 120  
DB 275 EYILGHLPATDIPQASRPLFKQNRVMTVDCHATPVRYQDESIMQPLCLGVSTLRAPH 334  
QY 121 GCHTQYMANMGVSAVLALAIIVYKGD-----SSKLMLGLVGHHCSPRYVPLRYAC 172  
DB 335 GCHAQYMANMGVSAVLITLAVITINGNDEAVGRSSRLMLGLVGHHTSARCLIPFLRYAC 394  
QY 173 EFLMQAFGLQMLQELQSLAEK 196  
DB 395 EFLMQAFGLQMLQELQSLSEK 418  
RESULT 5  
PHYB\_SORBI STANDARD: PRT, 1178 AA.  
AC P93527;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Phytochrome B.  
GN PHYB OR MA3.  
OS Sorghum bicolor (Sorghum) (Sorghum vulgare).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; liliopsida; Poales; Poaceae; PACC clade;  
OC Panicoideae; Andropogoneae; Sorghum.  
OX NCBI\_TaxID=4558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. 58M;  
RX MEDLINE=20188796; PubMed=10723737;  
RA Alba R., Kelmenson P.M., Cordonnier-Pratt M.-M., Pratt L.H.;  
RT "The phytochrome gene family in tomato and the rapid differential  
evolution of this family in angiosperms.";  
RL Mol. Biol. Evol. 17:362-373(2000).  
RN [2]  
RP SEQUENCE OF 208-1178 FROM N.A.  
RC STRAIN=CV. 58M;  
RX MEDLINE=97198556; PubMed=9046599;  
RA Childs K.L., Miller F.R., Cordonnier-Pratt M.-M., Pratt L.H.,  
RA Morgan P.W., Muller J.E.;  
RT "The sorghum bicolor photoperiod sensitivity gene, Ma3, encodes a  
phytochrome B.";  
RL Plant Physiol. 113:611-619(1997).  
CC -1- FUNCTION: REGULATORY PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT  
ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS  
MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT  
ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCONVERSION OF PR IN  
PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS  
RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE

CC RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR  
 CC GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-  
 CC BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,  
 CC PROTOCHLOROPHYLLIDE REDUCTASE, RNA, ETC. IT ALSO CONTROLS THE  
 CC EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION (BY  
 CC SIMILARITY).  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- PFM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.  
 CC -1- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.  
 CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, AF182394; AAA41398.2; -  
 DR InterPro: IPR003018; GAF.  
 DR InterPro: IPR003594; HATPase\_C.  
 DR InterPro: IPR004359; HIS\_KIN\_s1g.  
 DR InterPro: IPR003661; HIS\_kina.  
 DR InterPro: IPR000014; PAS.  
 DR InterPro: IPR001294; Phytochrome.  
 DR Pfam: PF01590; GAF; 1.  
 DR Pfam: PF02518; HATPase\_C; 1.  
 DR Pfam: PF00989; PAS; 2.  
 DR Pfam: PF00360; phytochrome; 1.  
 DR Pfam: PF00512; signal; 1.  
 DR PRINTS; PR01033; PHYTOCHROME.  
 DR SMART; SM00065; GAF; 1.  
 DR SMART; SM00387; HATPase\_C; 1.  
 DR SMART; SM00388; HISKA; 1.  
 DR SMART; SM00091; PAS; 2.  
 DR PROSITE; PS50109; HIS\_KIN; 1.  
 DR PROSITE; PS50112; PAS; 2.  
 DR PROSITE; PS00245; PHYTOCHROME\_1; 1.  
 DR PROSITE; PS50046; PHYTOCHROME\_2; 1.  
 KW Transcription regulation; Photoreceptor; Phytochrome; Chromophore;  
 KW Repeat; Multigene family.  
 FT DOMAIN 668 739 PAS 1.  
 FT DOMAIN 802 873 PAS 2.  
 FT DOMAIN 950 1170 HISTIDINE KINASE.  
 FT DOMAIN 23 31 POLY-HIS.  
 FT DOMAIN 43 54 POLY-GLY.  
 FT BINDING 372 372 CHROMOPHORE (BY SIMILARITY).  
 SO SEQUENCE 1178 AA; 129136 MW; C406DF21197B93F CRC64;

Query Match 78.7%; Score 806; DB 1; Length 1178;  
 Best Local Similarity 73.6%; Pred. No. 9.6e-74;  
 Matches 153; Conservative 21; Mismatches 22; Indels 12; Gaps 2;  
 QY 1 KLAVALAISRLDSLPGDGLGALCTVYEDVYQRLTGDRVMTYQFHEDDGEVSEIRSDL 60  
 DB 251 KLAVALAISRLDLPDGLDKLLCDPVEHEVRELTGYDRVAVYRFHDEDEGEVVAESRRDL 310  
 QY 61 EPLYGLHPADIDPOAARFLKONRVKICDGNAPVAVQVSEELKRLCLVNSLRAPH 120  
 DB 311 EPLYGLHPADIDPOASFLERQNRVRIADCHAPVAVVIDDPGNSQPLCLGSLRAH 370  
 QY 121 GCHTQYMANMGSAVALAIYVK--GKD-----SSKLMGVYGHCSPRVPEPL 168  
 DB 371 GCHQYMANMGSIASLVAVIISGGDBDQNGRGSISSAKMLGVLVCHHSPPRCPPL 430  
 QY 169 RYACEFLMQAFGLQLOMELQLASQIAEK 196  
 DB 431 RYACEFLMQAFGLQLMELQLAHQLSEK 458

RESULT 6  
 PHTB\_ARATH STANDARD: PRT; 1172 AA.  
 ID PHTB\_ARATH  
 AC P14713;  
 DT 01-APR-1990 (Rel. 14, Created).  
 DF 01-APR-1990 (Rel. 14, Last sequence update)  
 DE 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Phytochrome B.  
 GN PHTB OR HY3 OR AT2G18790 OR MSF3.17.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eustoids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=90108670; PubMed=2606345;  
 RA Sharrock R.A., Quail P.H.;  
 RT "Novel phytochrome sequences in Arabidopsis thaliana: structure,  
 RT evolution, and differential expression of a plant regulatory  
 RT photoreceptor family.";  
 RL Genes Dev. 3:1745-1757(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. LANDSBERG ERRECTA;  
 RX MEDLINE=93200802; PubMed=8453299;  
 RA Reed J.W., Nagpal P., Poole D.S., Furuya M., Chory J.;  
 RT "Mutations in the gene for the red/far-red light receptor phytochrome  
 RT B alter cell elongation and physiological responses throughout  
 RL Arabidopsis development.";  
 RL Plant Cell 5:147-157(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-T., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.D., Ronning C.M., Koo H.L.,  
 RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,  
 RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Cressy T.H.,  
 RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,  
 RA Nieman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,  
 RA Venter J.C.;  
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 RT thaliana.";  
 RL Nature 402:761-768(1999).  
 CC -1- FUNCTION: REGULATOR PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT  
 CC ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS  
 CC MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT  
 CC ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCOVERSION OF PR IN  
 CC PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS  
 CC RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE  
 CC RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR  
 CC GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-  
 CC BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,  
 CC PROTOCHLOROPHYLLIDE REDUCTASE, RNA, ETC. IT ALSO CONTROLS  
 CC THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- PFM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.  
 CC -1- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.  
 CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, X17342; CAA35222.1; -

DR EMBL: L09262; -: NOT ANNOTATED\_CDS.  
DR PIR: S07718; FKMOB.  
DR PIR: JQ2141; JQ2141.  
DR Mendel: 1310; ARACH.PhyB.1.  
DR InterPro: IPR003018; GAF.  
DR InterPro: IPR003594; HATPase\_C.  
DR InterPro: IPR004359; HIS\_KIN\_S19.  
DR InterPro: IPR003661; HIS\_KIN.  
DR InterPro: IPR000014; PAS.  
DR InterPro: IPR001294; Phytochrome.  
DR Pfam: PF01590; GAF; 1.  
DR Pfam: PF02518; HATPase\_C; 1.  
DR Pfam: PF00989; PAS; 2.  
DR Pfam: PF00360; Phytochrome; 1.  
DR Pfam: PF00512; signal; 1.  
DR PRINTS: PRO1033; PHYTOCHROME.  
DR SMART: SM00065; GAF; 1.  
DR SMART: SM00387; HATPase\_C; 1.  
DR SMART: SM00388; HSKA; 1.  
DR SMART: SM00091; PAS; 2.  
DR PROSITE: PS50109; HIS\_KIN; 1.  
DR PROSITE: PS50112; PAS; 2.  
DR PROSITE: PS00245; PHYTOCHROME\_1; 1.  
DR PROSITE: PS50046; PHYTOCHROME\_2; 1.  
KW Transcription regulation; Photoreceptor; Phytochrome; Chromophore;  
KW Repeat; Multigene family.  
FT DOMAIN 652 723 PAS 1.  
FT DOMAIN 786 857 HISTIDINE KINASE.  
FT DOMAIN 934 1153 GLY/SER-RICH.  
FT BINDING 357 357 CHROMOPHORE.  
SQ SEQUENCE 1172 AA; 129331 MW; 7B5348CB1091B813 CRC64;

Query Match 78.4%; Score 802.5; DB 1; Length 1172;  
Best Local Similarity 72.9%; Pred. No. 2.2e-73;  
Matches 151; Conservative 23; Mismatches 22; Indels 11; Gaps 1;

QY 1 KLAVRAISRLQSLPGDGLGALCDTVVEYQRLTGDRVYVQFHHDDHGEVYSEIRSDL 60  
DB 236 KLAVRAISRLQSLPGDGLGALCDTVVEYQRLTGDRVYVQFHHDDHGEVYSEIRSDL 295  
QY 61 EPYGLNHPATDIPQARFLFKONRVRMICDNPVAVVQSEELKRLCLVNSTLRAPH 120  
DB 296 EPYGLNHPATDIPQARFLFKONRVRMICDNPVAVVQSEELKRLCLVNSTLRAPH 355  
QY 121 GCHTOYANMNGSVASLALAIYV-----KGKDSKIMGLGVGHHCSPRYVPEPLR 169  
DB 356 GCHTOYANMNGSVASLALAIYV-----KGKDSKIMGLGVGHHCSPRYVPEPLR 415  
QY 170 YACEFLMQAEGLOLQMLQSLAQAEK 196  
DB 416 YACEFLMQAEGLOLQMLQSLAQAEK 442

RESULT 7  
PRTB\_ORYSA STANDARD; PRT; 1171 AA.  
ID PRTB\_ORYSA  
AC P25764;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Phytochrome B.  
GN PHYB OR PHYB1.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. INDICA-IR36; TISSUE=Seedling shoot;

RX MEDLINE-91172131; PubMed-2005872;  
RA Denesh K., Tepperman J., Christensen A.H., Quail P.H.;  
RT "PhyB is evolutionarily conserved and constitutively expressed in  
rice seedling shoots.";  
RL Mol. Gen. Genet. 225:305-313(1991).  
CC -!- FUNCTION: REGULATORY PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT  
CC ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT. THE PR FORM THAT ABSORBS  
CC MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT  
CC ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCOVERSION OF PR IN  
CC PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS  
CC RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE  
CC RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR  
CC GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-  
CC BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,  
CC PROTOCHLOROPHYLLIDE REDUCTASE, RNA, ETC. IT ALSO CONTROLS  
CC THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION.  
CC -!- SUBUNIT: HOMODIMER.  
CC -!- PMT: CONTAINS ONE COVALENTLY LINKED TETRAPPTROLE CHROMOPHORE.  
CC -!- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.  
CC -!- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
CC -!- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb.ch](mailto:license@isb.ch)).  
CC -----  
DR EMBL: X57563; CAA40795.2; -;  
DR PIR: S14065; S14065.  
DR Mendel: 1319; ORYSA.PhyB.1.  
DR InterPro: IPR003018; GAF.  
DR InterPro: IPR003594; HATPase\_C.  
DR InterPro: IPR004359; HIS\_KIN\_S19.  
DR InterPro: IPR003661; HIS\_KIN.  
DR InterPro: IPR000014; PAS.  
DR InterPro: IPR001294; Phytochrome.  
DR Pfam: PF01590; GAF; 1.  
DR Pfam: PF02518; HATPase\_C; 1.  
DR Pfam: PF00989; PAS; 2.  
DR Pfam: PF00360; Phytochrome; 1.  
DR PRINTS: PRO1033; PHYTOCHROME.  
DR SMART: SM00065; GAF; 1.  
DR SMART: SM00387; HATPase\_C; 1.  
DR SMART: SM00388; HSKA; 1.  
DR SMART: SM00091; PAS; 2.  
DR PROSITE: PS50109; HIS\_KIN; 1.  
DR PROSITE: PS50112; PAS; 2.  
DR PROSITE: PS00245; PHYTOCHROME\_1; 1.  
DR PROSITE: PS50046; PHYTOCHROME\_2; 1.  
KW Transcription regulation; Photoreceptor; Phytochrome; Chromophore;  
KW Repeat; Multigene family.  
FT DOMAIN 661 732 PAS 1.  
FT DOMAIN 795 866 PAS 2.  
FT DOMAIN 943 1161 HISTIDINE KINASE.  
FT DOMAIN 39 51 POLY-GLY.  
FT BINDING 364 364 CHROMOPHORE (BY SIMILARITY).  
SQ SEQUENCE 1171 AA; 128384 MW; E8292E88B769BF16 CRC64;

Query Match 77.6%; Score 794.5; DB 1; Length 1171;  
Best Local Similarity 71.8%; Pred. No. 1.4e-72;  
Matches 150; Conservative 23; Mismatches 23; Indels 13; Gaps 2;

QY 1 KLAVRAISRLQSLPGDGLGALCDTVVEYQRLTGDRVYVQFHHDDHGEVYSEIRSDL 60  
DB 243 KLAVRAISRLQSLPGDGLGALCDTVVEYQRLTGDRVYVQFHHDDHGEVYSEIRSDL 302  
QY 61 EPYGLNHPATDIPQARFLFKONRVRMICDNPVAVVQSEELKRLCLVNSTLRAPH 120





CC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.  
 OX NCBI\_TaxID=3847;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=CV. PALDAR; TISSUE=Etolated leaf;  
 RA Hahn T.R., Woo T.W., Seo H.S., Choi Y.D.;  
 RL Submitted (xxx-1994) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: REGULATOR PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT  
 CC ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS  
 CC MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PER FORM THAT  
 CC ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCONVERSION OF PR IN  
 CC PER INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS  
 CC RECONVERSION OF PER TO PR CANCELS THE INDUCTION OF THOSE  
 CC RESPONSES. PER CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR  
 CC GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-  
 CC BIPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,  
 CC PROTOCHLOROPHYLLIDE REDUCTASE, RNA, ETC. IT ALSO CONTROLS  
 CC THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- PTM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.  
 CC -1- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L34843; AAA34000.1; -  
 DR Mendel; 8366; GLYMA.PHYB.1;  
 DR InterPro: IPR003018; GAF.  
 DR InterPro: IPR003594; HATPase\_c.  
 DR InterPro: IPR004359; HIS\_KIN\_sig.  
 DR InterPro: IPR003661; HIS\_kina.  
 DR InterPro: IPR000014; PAS.  
 DR InterPro: IPR001294; Phytochrome.  
 DR Pfam: PF01590; GAF; 1.  
 DR Pfam: PF02518; HATPase\_c; 1.  
 DR Pfam: PF00989; PAS; 2.  
 DR Pfam: PF00360; phytochrome; 1.  
 DR Pfam: PF00512; signal; 1.  
 DR PRINTS: PR01033; PHYTOCHROME.  
 DR SMART: SM00065; GAF; 1.  
 DR SMART: SM00387; HATPase\_c; 1.  
 DR SMART: SM00388; HISKA; 1.  
 DR SMART: SM00091; PAS; 1.  
 DR PROSITE: PS50109; HIS\_KIN; 1.  
 DR PROSITE: PS50112; PAS; 1.  
 DR PROSITE: PS00245; PHYTOCHROME\_1; 1.  
 DR PROSITE: PS50046; PHYTOCHROME\_2; 1.  
 KW Transcription regulation; Photoreceptor; Phytochrome; Chromophore;  
 KW Multigene family.  
 FT DOMAIN 646 717 PAS.  
 FT BINDING 358 358 HISTIDINE KINASE.  
 FT BINDING 358 358 CHROMOPHORE (BY SIMILARITY).  
 SO SEQUENCE 1156 AA; 129085 MW; 76333AABDC42D297 CRC64;

Query Match 75.4%; Score 772; DB 1; Length 1156;  
 Best Local Similarity 72.6%; Pred. No. 2.7e-70;  
 Matches 146; Conservative 20; Mismatches 27; Indels 8; Gaps 1;

OY 4 VRAISRLOSLEGDICALCDYVENVQRLTGYDRWYQFHHDDHGEVYSERSDLEPY 63  
 DB 240 VRAISQLOSLEPSADYKLCIDYVESVRELGTGDRWYVKFHEHDEGEVYSESKRDLDEPY 299  
 OY 64 LGHHPATDIOQAARFLKONRVRMIGCCNATPVYVVOSEELKRPICLVNSTRPHGH 123  
 DB 300 IGLHPTDIDFQASRFLKONRVRMIVDCIHASAVVODEALVOPICLVGSLTGAIPHGH 359

OY 124 TOYMANMGSVASLALAIYVKGD-----SSKLMGLVGHHCSPRYVPPPLRYACEFL 175  
 DB 360 AQYMANMGSIASLVAAVIINGNDEGVGRSSMRMLGLVCHHTSARCIPLFLRYACEFL 419  
 OY 176 MOAFGLQIMELQIASQAEK 196  
 DB 420 MOAFGLQIMELQIAASLEK 440

RESULT 10  
 PHY\_PICAB  
 ID PHY\_PICAB STANDARD; PRT; 1136 AA.  
 AC Q40762;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Phytochrome.  
 OS Picea abies (Norway spruce) (Picea excelsa).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.  
 OX NCBI\_TaxID=3329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Clapham D.H., Larsson C.T., Qamaruddin M.;  
 RL Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: REGULATOR PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT  
 CC ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS  
 CC MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PER FORM THAT  
 CC ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCONVERSION OF PR IN  
 CC PER INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS  
 CC RECONVERSION OF PER TO PR CANCELS THE INDUCTION OF THOSE  
 CC RESPONSES. PER CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR  
 CC GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-  
 CC BIPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,  
 CC PROTOCHLOROPHYLLIDE REDUCTASE, RNA, ETC. IT ALSO CONTROLS  
 CC THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION.  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- PTM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.  
 CC -1- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.  
 CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U60264; AAB03339.1; -  
 DR InterPro: IPR003018; GAF.  
 DR InterPro: IPR003594; HATPase\_c.  
 DR InterPro: IPR004359; HIS\_KIN\_sig.  
 DR InterPro: IPR003661; HIS\_kina.  
 DR InterPro: IPR000014; PAS.  
 DR InterPro: IPR001294; Phytochrome.  
 DR Pfam: PF01590; GAF; 1.  
 DR Pfam: PF02518; HATPase\_c; 1.  
 DR Pfam: PF00989; PAS; 2.  
 DR Pfam: PF00360; phytochrome; 1.  
 DR Pfam: PF00512; signal; 1.  
 DR PRINTS: PR01033; PHYTOCHROME.  
 DR SMART: SM00065; GAF; 1.  
 DR SMART: SM00387; HATPase\_c; 1.  
 DR SMART: SM00388; HISKA; 1.  
 DR SMART: SM00091; PAS; 2.  
 DR PROSITE: PS50109; HIS\_KIN; 1.  
 DR PROSITE: PS50112; PAS; 2.  
 DR PROSITE: PS00245; PHYTOCHROME\_1; 1.  
 DR PROSITE: PS50046; PHYTOCHROME\_2; 1.  
 KW Transcription regulation; Photoreceptor; Phytochrome; Chromophore;





RL Plant Mol. Biol. 20:589-600(1992).

CC -1- FUNCTION: REGULATOR PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT

CC ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS

CC MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PER FORM THAT

CC ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCONVERSION OF PR IN

CC PER INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS

CC RECONVERSION OF PER TO PR CANCELS THE INDUCTION OF THOSE

CC RESPONSES. PER CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR

CC GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-

CC BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,

CC PHOTOCOLOROPHYLLIDE REDUCTASE, RNA, ETC. IT ALSO CONTROLS

CC THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION.

CC -1- SUBUNIT: HOMODIMER.

CC -1- PTM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.

CC -1- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.

CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.

CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----

CC EMBL: S51538; AAB24397.1; -.

CC PIR: S28431; S28431.

CC Mendel: 10718; SOLtu:PHYB.1.

CC InterPro: IPR003018; GAF.

CC InterPro: IPR003594; HATPase\_C.

CC InterPro: IPR004359; HIS\_KIN\_Sig.

CC InterPro: IPR003661; HIS\_KIN.

CC InterPro: IPR001294; Phytochrome.

CC Pfam: PF01590; GAF. 1.

CC Pfam: PF02518; HATPase\_C. 1.

CC Pfam: PF00989; PAS. 2.

CC Pfam: PF00360; phytochrome. 1.

CC Pfam: PF00512; signal. 1.

CC PRINTS: PRO1033; PHYTOCHROME.

CC SMART: SM00065; GAF. 1.

CC SMART: SM00387; HATPase\_C. 1.

CC SMART: SM00388; HISKA. 1.

CC SMART: SM00091; PAS. 2.

CC PROSITE: PS50109; HIS\_KIN. 1.

CC PROSITE: PS50112; PAS. 2.

CC PROSITE: PS00245; PHYTOCHROME\_1; 1.

CC PROSITE: PS50046; PHYTOCHROME\_2; 1.

CC KEGG: Transcription regulation; Photoreceptor; Phytochrome; Chromophore;

CC Repeat: Multigene family.

CC KEGG: Repeat: Multigene family.

CC FT DOMAIN 622 693 PAS 1.

CC FT DOMAIN 755 826 PAS 2.

CC FT DOMAIN 903 1122 HISTIDINE KINASE.

CC FT BINDING 333 333 CHROMOPHORE.

CC SQ SEQUENCE 1129 AA; 125621 MW; 60DF5FB964EFC60B CRC64;

QY Query Match 74.6%; Score 763.5; DB 1; Length 1129;

DB Best Local Similarity 74.7%; Pred. No. 1.9e-69;

Matches 142; Conservative 18; Mismatches 21; Indels 9; Gaps 1;

QY 16 GDIGALCDTVVEDYQRLTGYDRVMVYGFPHEDDGEVSEIRKSDLEPLGLHYPATDIPQ 75

DB 227 GTLKLCCTVVESEYELGYDRVMVYKFEHDEHGEVSEKRSKSDLEPLGLHYPATDIPQ 286

QY 76 AARLFKONRVRMTCDCNATPVKYVQSEELKRLPLCLVNSTLRAPGCHTGYMANNGSVAS 135

DB 287 ASRFLFKONRVRMIVDCHATFVRVYQDESMLQPLCLVNSTLRAPGCHTGYMANNGSVAS 346

QY 136 LALAIIVVGRD-----SSKLMGLVVGHHGCSRRVYRPFPLRYACETLMQAFGLQLOME 186

DB 347 LTLAVIINGNDEAVGGRNSMRMLGLVVGHHGTSVRSIPFLRYACEFLMQAFGLQLOME 406

QY 187 LQLASQLAEK 196

DB 407 LQLASQLSEK 416

RESULT 13

PHY\_ADICA

ID PHY\_ADICA STANDARD; PRT; 1118 AA.

AC P42496;

DT 30-NOV-1995 (Rel. 32, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Phytochrome 1.

GN PHY1.

OS Adiantum capillus-veneris (Fern).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Filicophyta; Filicopsida; Filicales; Adiantaceae; Adiantum.

OX NCBI\_TaxID=13818;

RM [1]

RP SEQUENCE FROM N.A.

RA Okamoto H., Hirano Y., Abe H., Tomizawa K.I., Furuya M., Wada M.;

RT "The deduced amino sequence of phytochrome from *Adiantum* includes

RL consensus motifs present in phytochrome B from seed plants.";

PL Plant Cell Physiol. 34:1329-1334(1993).

RM [2]

RP REVISIONS TO 28; 93-97; 310; 345 AND 1044.

RA Nozue K., Fukuda S., Kanegae T., Wada M.;

RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.

CC -1- FUNCTION: REGULATOR PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT

CC ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS

CC MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PER FORM THAT

CC ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCONVERSION OF PR IN

CC PER INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS

CC RECONVERSION OF PER TO PR CANCELS THE INDUCTION OF THOSE

CC RESPONSES. PER CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR

CC GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-

CC BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,

CC PHOTOCOLOROPHYLLIDE REDUCTASE, RNA, ETC. IT ALSO CONTROLS

CC THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION.

CC -1- SUBUNIT: HOMODIMER.

CC -1- PTM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.

CC -1- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.

CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.

CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----

CC EMBL: AB016168; BAA51856.1; -.

CC InterPro: IPR003018; GAF.

CC InterPro: IPR003594; HATPase\_C.

CC InterPro: IPR004359; HIS\_KIN\_Sig.

CC InterPro: IPR003661; HIS\_KIN.

CC InterPro: IPR001610; PAC.

CC InterPro: IPR000014; PAS.

CC InterPro: IPR001294; Phytochrome.

CC Pfam: PF01590; GAF. 1.

CC Pfam: PF02518; HATPase\_C. 1.

CC Pfam: PF00989; PAS. 2.

CC Pfam: PF00360; phytochrome. 1.

CC Pfam: PF00512; signal. 1.

CC PRINTS: PRO1033; PHYTOCHROME.

CC SMART: SM00065; GAF. 1.

CC SMART: SM00387; HATPase\_C. 1.

CC SMART: SM00388; HISKA. 1.

CC SMART: SM00086; PAC. 1.

**DR SMART:** SMO0091; PAS: 2.

**DR PROSITE;** PS50109; HIS\_KIN: 1.

**DR PROSITE;** PS50112; PAS: 2.

**DR PROSITE;** PS00245; PHYTOCHROME\_1; 1.

**DR PROSITE;** PS50046; PHYTOCHROME\_2; 1.

**KW** Transcription regulation; Photoreceptor; Phytochrome; Chromophore;  
Repeat.

<b>FT DOMAIN</b>	606	677	PAS 1.
<b>FT DOMAIN</b>	740	811	PAS 2.
<b>FT BINDING</b>	887	1110	HISTIDINE KINASE.
<b>SQ SEQUENCE</b>	1118 AA;	123824 MW;	27C21BF61FAB9333 CRC64;

  

**Query Match**           **74.1%;** Score 758.5; DB 1; Length 1118;  
**Best Local Similarity** 67.8%; Pred. No. 6e-65;  
**Matches 139;** Conservative 26; Mismatches 31; Indels 9; Gaps

---

```

Oy      1 KLAVAIARLSLPGCDIGALCDTVEVDVQRLLTGVDYRWVVOGFHEDDHEEVSSEIRSDL 60
Db      196 KLAQAIAIKRSLSPGCGDIGLCDSVVEEVHELTLGDRWAYAKFFHDEHGEEVAIEIRRTL 255
Oy      61 EPLYGLHPATDIPOAAREFLFKONFRVMICDCNAPRVKVGQSSEELRPCLVLNSTLRADP 120
Db     256 EPYLGHYPATDIPOAAAEFLFMKNENVRMIDCRLEPPVKLIQDTKSQPMSLTSGSKLRAPH 315
Oy     121 GCRTGYMANMGSAVALATAIVYGKDSS-----KLNGIYVGHGSRRVPPEPLARA 171
Db     316 GCHTIHYMMNMSTISLVMAVIYNDDSDDSPGSHSOGIKLMGLVCVGHHSPRVVPEPVNSA 375
Oy     172 CEFLMQAFGLQLOMELQLASAQEAER 196
Db    376 CEFLMQVFSLQIMNEGVMAAGVREKR 400

RESULT   14
PHYL_SEIMA PHYL_SEIMA STANDARD: PRT: 1134 AA.
AC Q00549;
AD 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE phytochrome 1.
GN PHYL1.
OS Selaginella martensii (Martens's spike moss).
OC Eukaryota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Lycopodiophyta; Isoetopsida; Selaginellales; Selaginellaceae; Selaginella NCBI_TaxID=3247;
OX [1]
RN RP SEQUENCE FROM N.A. STRAIN-SPRING; MEDLINE=93117303; PubMed=1475321; RA Hanelt S., Braun B., Marx S., Schneider-Poetsch H.A.W.; RT "phytochrome evolution: a phylogenetic tree with the first complete sequence of phytochrome from a cryptogamic plant (Selaginella martensi) spring"; RL Photochem. Photobiol. 56:751-758(1992).
CC - FUNCTION: REGULATOR PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT. THE PR FORM THAT ABSORB MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PER FORM THAT ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCOVERSION OF PR IN PER INDUCES AN ARRAY OF MORPHOGENIC RESPONSES. WHEREAS RECONVERSION OF PER TO PR CANCELS THE INDUCTION OF THESE RESPONSES. PER CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-BIPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN, PROTHIOESTROPYLIDE REDUCTASE, RENNA, ETC. IT ALSO CONTROLS THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION. CC -1- SUBUNIT: HOMODIMER.
CC -1- PTR: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.
CC -1- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.
CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARN-T-SIM) DIMERIZATION DOMAINS.

```

[illegible]

DT 01-JUN-1994 (rel. 29, Created)  
 DT 01-JUN-1994 (rel. 29, Last sequence update)  
 DT 01-MAR-2002 (rel. 41, Last annotation update)  
 DE Phytochrome 1.  
 GN PHY1.  
 OS Physcomitrella patens (Moss).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;  
 OC Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.  
 OX NCBI\_TaxId=3218;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94039823; PubMed=8224238;  
 RA Kolukisaoglu H.U., Braun B., Martin W.F., Schneider-Poetsch H.A.W.;  
 RT "Mosses do express conventional, distantly B-type-related  
 phytochromes. Phytochrome of Physcomitrella patens (Hedw.).";  
 RL FEBS Lett. 334:95-100(1993).  
 CC -1- FUNCTION: REGULATORY PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT  
 ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS  
 MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT  
 ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCONVERSION OF PR IN  
 PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS  
 RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE  
 RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR  
 GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-  
 BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,  
 PROTOCHLOROPHYLLIDE REDUCTASE, RNA, ETC. IT ALSO CONTROLS  
 THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- PTM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.  
 CC -1- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.  
 CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See: <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X75025; CAA52933.1; -  
 DR PIR: S37206; S37206.  
 DR InterPro: IPR003018; GAF.  
 DR InterPro: IPR003594; HATPase\_C.  
 DR InterPro: IPR004359; HIS\_KIN\_sig.  
 DR InterPro: IPR003661; HIS\_KIN\_A.  
 DR InterPro: IPR001610; PAC.  
 DR InterPro: IPR000014; PAS.  
 DR InterPro: IPR001294; Phytochrome.  
 DR Pfam: PF01590; GAF; 1.  
 DR Pfam: PF02518; HATPase\_C; 1.  
 DR Pfam: PF00989; PAS; 2.  
 DR Pfam: PF00360; phytochrome; 1.  
 DR Pfam: PF00512; signal; 1.  
 DR PRINTS: PRO1033; PHYTOCHROME.  
 DR SMART: SM00065; GAF; 1.  
 DR SMART: SM00387; HATPase\_C; 1.  
 DR SMART: SM00388; HISKA; 1.  
 DR SMART: SM00086; PAC; 1.  
 DR SMART: SM00091; PAS; 2.  
 DR PROSITE: PS50109; HIS\_KIN; 1.  
 DR PROSITE: PS50112; PAS; 2.  
 DR PROSITE: PS00245; PHYTOCHROME\_1; 1.  
 DR PROSITE: PS50046; PHYTOCHROME\_2; 1.  
 KW Transcription regulation; Photoreceptor; Phytochrome; Chromophore;  
 KW Repeat.  
 FT DOMAIN 610 681 PAS 1.  
 FT DOMAIN 744 815 PAS 2.  
 FT DOMAIN 895 1115 HISTIDINE KINASE.  
 FT BINDING 321 321 CHROMOPHORE (BY SIMILARITY).  
 SQ SEQUENCE 1132 AA; 125230 MW; E1DAD4D6DC9CD16 CRC64;

Query Match 72.3%; Score 740.5; DB 1; Length 1132;  
 Best Local Similarity 66.3%; Pred. No. 4,1e-67;  
 Matches 136; Conservative 29; Mismatches 31; Indels 9; Gaps 1;  
 QY 1 KLAIVRAISRLQSLPGDIGALCDTVYEDVQRLTGIDRWVYQGFHEDDHEVSEIRSDL 60  
 DB 200 KLAAKAIFRIQALPGGNIGILCDTVVEEVEVELTGIDRWVAYRFEHDEHGEVVAEIRRADL 259  
 QY 61 EPTLGLHYPATDIPQARFLFKONRYRMICDGNATPQVQSEELKRPCLVNSTLRAP 120  
 DB 260 EPTLGLHYPTDIPQASRFLFKNNKRYITADSAPPVKYIQDPTLRQVSLAGSTIRSH 319  
 QY 121 GCHTYQYMANNGSVASIALAIIVYKGDSS-----KIMGIVGHHCSPPRYVPELRYA 171  
 DB 320 GCHAYGMNGMSIASLVMAVITINDNEDSHGVSQGRKLMGLVCHHTSPRTVPPFLRSA 379  
 QY 172 CEFLMQAFGLQLOMELQLASQLAEK 196  
 DB 380 CGFLMQVFGLOLMEVESAAQLREK 404

Search completed: June 7, 2002, 18:58:04  
 Job time: 243 sec